THE EVOLUTIONARY HISTORY AND GENOMIC IMPACT OF MAMMALIAN DNA TRANSPOSONS

by

JOHN KELLY PACE, II

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2008

ACKNOWLEDGEMENTS

I would like to first and foremost thank my wonderful wife, Deedra, and my incredible children Jana, Julie and Jill, for their tireless support during graduate school. They have spent many hours without me while I was working on the computer, but their faith in me and willingness to allow me to follow my dream has kept me going when I was ready to stop. I love you all and cannot say thank you enough!

Next, I want to thank my advisor, Dr. Cedric Feschotte. Cedric has not only been a great advisor, but has also been a good friend. He has provided me with opportunities to grow intellectually and given me chances to learn from others in places I would never have been able to go without his help such as France. He allowed me to be a Research Assistant in his lab, thereby giving me more time to focus on my research and finish school in a timely manner. I am forever grateful for all he has done for me.

I would like to thank the Biology Department at the University of Texas at Arlington. My committee members, Drs. Esther Betran, Shawn Christensen, Andre Pires da Silva and Ellen Pritham, who challenged me to learn more than I thought was possible, knowing full well it would make me a better scientist. The members of the Feschotte Lab have been a tremendous support, particularly Don Hucks, Marcel Guibotsy, Claudio Casola, Qi Wang and Clement Gilbert. It has been a privilege to work alongside each of them. I would like to thank the members of the Genome Biology Group for their insights and comments on my work and for allowing me to learn from their work. My work on human polymorphism of repair-mediated duplications would not have been possible without the assistance of Mark Batzer and Shurjo Sen at Louisiana State University. Last, I would like to thank the Biology Office staff, Gloria Burlingham, Linda Taylor and Peggy Fisher for all of their help.

October 15, 2008

ABSTRACT

THE EVOLUTIONARY HISTORY AND GENOMIC IMPACT OF MAMMALIAN DNA TRANSPOSONS

John Kelly Pace, PhD.

The University of Texas at Arlington, 2008

Supervising Professor: Cedric Feschotte

Transposable elements (TEs) are mobile repetitive sequences that make up large fractions of mammalian genomes, including at least 45% of the human genome. Class 2, or DNA intermediate transposons, which make up approximately 3% of human nuclear DNA, have typically been overlooked in analyses of mammalian genomes and therefore remain poorly understood. Here, we carried out the first large-scale analysis of the evolutionary history of DNA transposons throughout eutherian mammal evolution. The analysis included 18 mammals from all major branches of the eutherian tree including Afrotheria, Laurasiatheria, primates, rodents and Xenarthra. We combined three different computational methods: average divergence of the TE family from the ancestral consensus sequence, nested insertion analysis and cross-species genomic analysis of orthologous loci to determine the average age of each TE family, when the family was active and in which species. The combination of these methods, the latter two of which do not rely upon calibration of a molecular clock, allowed us to trace the evolutionary history of the 249 currently recognizable eutherian DNA transposon families for which at least 100 copies of the TE could be identified.

Our analysis revealed that, contrary to previous assumptions, the horizontal transfer (HT) of DNA transposons is a widespread and common phenomenon in eutherians. We report strong evidence for the HT of 13 different autonomous TE families that were horizontally transferred into 13 of the 18 eutherian species surveyed, typically invading multiple species lineages. In each case, the infiltration of the TE family was an independent event that resulted in lineage-specific activity, producing distinct bursts of transposition. Together, these bursts were responsible for the insertion of between 2,300 and 222,000 copies of new elements and the addition of up to 39 megabases of nuclear DNA per species. We discovered a general, eutherian-wide slowdown in the number of horizontal transfers and lineage-specific activity of DNA transposons over the past 100 million years, with only 2 different families of TEs invading 4 of the 18 species within the past 40 million years.

TABLE OF CONTENTS

APPENDIX

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

THE EVOLUTIONARY HISTORY OF HUMAN DNA TRANSPOSONS: EVIDENCE FOR INTENSE ACTIVITY IN THE PRIMATE LINEAGE

1.1 Introduction

Transposable elements (TEs) are mobile repetitive sequences that make up large fractions of mammalian genomes, including at least 45% of the human genome (Lander et al. 2001), 37.5% of the mouse genome (Waterston et al. 2002), and 41% of the dog genome (Lindblad-Toh et al. 2005). TEs may be classified based upon their method of transposition. Class 1 elements transpose via an RNA intermediate using reverse transcriptase and include long and short interspersed nuclear elements (LINEs and SINEs), and long terminal repeats elements. Class 2 elements, or DNA transposons, transpose via a DNA intermediate through a so-called cut-and-paste mechanism (Craig et al. 2002).

Information on human DNA transposons is currently very scarce. This type of element makes up 3% of our genome (Lander et al. 2001), yet only a limited number of studies have focused on DNA transposons in any mammalian genomes (Auge-Gouillou et al. 1995; Morgan 1995; Oosumi et al. 1995; Robertson 1996; Smit and Riggs 1996; Robertson and Martos 1997; Robertson and Zumpano 1997; Dematteï et al. 2000). In contrast, the evolutionary history and genomic impact of mammalian retrotransposons has been the subject of intensive investigation (Smit et al. 1995; Kapitonov and Jurka 1996; Szak et al. 2002; Xing et al. 2003; Price et al. 2004; Khan et al. 2006; Deininger et al. 2003). This gap in knowledge can largely be explained by the relatively recent discovery of DNA transposons. Just a decade ago, several groups independently reported the presence of two different families of *mariner*-like elements in the human genome (Morgan 1995; Oosumi et al. 1995; Smit and Riggs 1996), now called Hsmar1 and Hsmar2. The evolutionary history of these two families was analyzed in detail by

Robertson and colleagues using the genomic sequence data available at that time. These studies indicated that Hsmar1 was active during early primate evolution, about 50 My ago (Robertson and Zumpano 1997), while Hsmar2 was older, having propagated at least 80 My ago (Robertson and Martos 1997).

In a seminal study dedicated to human DNA transposons, Smit & Riggs (1996) estimated that over 150,000 nonautonomous miniature inverted-repeat transposable elements (MITEs) from three major evolutionary groups, hAT, pogo and *mariner*, were integrated in the human genome. In addition to these MITEs, multiple lineages of elements with coding capacity for seemingly full-length but corrupted transposase were identified in the human genome (Morgan 1995; Smit and Riggs 1996; Smit 1999). More recently, members of other eukaryotic superfamilies have been identified in the human genome, including the piggyBac, Merlin and Mutator superfamilies (Smit 1999; Sarkar et al. 2003; Feschotte 2004; Jurka et al. 2005) as well as single-copy genes derived from transposases of the P-element and PIF/Harbinger superfamilies (Hagemann and Pinsker 2001; Kapitonov and Jurka 2004; Zhang et al. 2004). Overall, 7 out of 9 known eukaryotic superfamilies of DNA transposons are represented in the human genome and 125 different families are currently listed in Repbase Update (Jurka et al. 2005) that have a copy number of 100 or greater. Only a handful of these families have been subject to a detailed analysis.

The most comprehensive age analysis of human DNA transposons published to date appeared in the initial analysis of the human genome sequence. This study concluded that "there is no evidence for DNA transposon activity in the past 50 My in the human genome" (Lander et al. 2001). However, this statement should be taken with caution, since this conclusion was drawn solely from the average level of nucleotide divergence of individual copies to their reconstructed family consensus sequence. This approach has proven generally reliable to date the most prominent bursts of TE amplification and for comparing groups of elements that evolve at the same or almost the same rate. However, due to rapid fluctuations in substitution rates during the mammalian and primate radiation (Goodman 1985; Yi et al. 2002)

and possible variations in the substitution rate of different types of TEs (Xing et al. 2004), the results of these analyses should be interpreted carefully. In addition, the method is entirely dependent on the reconstruction of an accurate consensus sequence, a process that is sensitive to the number and genomic distribution of the elements. Since human DNA transposons are anticipated to be of relatively ancient origin and have likely diversified over the entire course of mammalian evolution, sequence divergence should not be the sole method used for accurately dating these types of elements. Furthermore, there has been no published effort to characterize the tempo and mode of accumulation of every DNA transposon family in any mammalian genome. Therefore an accurate and detailed picture of DNA transposon history in humans and other mammals is still lacking.

Here we present the first detailed analysis of the age of the 125 DNA transposon families currently recognizable in the human genome. In particular, we sought to evaluate which human DNA transposon families were actively transposing during primate evolution. To this end, we employed a combination of three independent computational methods, two of which do not rely upon sequence divergence. We estimate that at least 40 families of DNA transposons were active during the primate radiation. We conclude that ~98,000 individual elements were added to the primate genome in the last ~80 My of evolution. Eleven of these families, or ~23,000 individual elements, inserted into the primate genome between the split of prosimian primates and new world monkeys (~40 to 63 Mya). However, we found no evidence that any human DNA transposon family was active within the last ~37 My of primate evolution. Our results suggest an intriguing history of intense activity of diverse DNA transposons during the first half of the primate radiation, followed by a striking cessation of transposition activity in an anthropoid primate ancestor and no detectable germ-line reinfiltration of the primate lineages leading to humans over the last 37 Myr.

1.2 Results

1.2.1 Census of DNA transposons in the human genome

We began our investigation by assessing the diversity and copy number of all DNA transposon families currently recognized in the human genome. Copy numbers were calculated from the RepeatMasker tables of the May 2004 assembly of the human genome, available through the UCSC Genome Browser (http://genome.ucsc.edu). In agreement with previous reports (Smit and Riggs 1996; Smit 1999; Lander et al. 2001), we found that two superfamilies, hAT and Tc1/*mariner*, are predominant in the human DNA transposon population (Table 1.1). hAT elements account for approximately two thirds of the census and more than half of the 125 families. Human Tc1/*mariner* elements account for one-third of the population and can be divided into three evolutionary distinct lineages: pogo-like, *mariner*-like and Tc2-like (Smit and Riggs 1996; Robertson 2002). The former is the most abundant and diversified lineage, and includes 8 families of transposase-encoding Tigger elements and 22 related MITE families. The prevalence of nonautonomous MITEs (74% of the total number of DNA TEs) over transposase-encoding elements (26%) is particularly striking in the human genome and this phenomenon affects all superfamilies (Table 1.1). It is also a characteristic of the DNA transposon population of plants and nematodes (Feschotte et al. 2002).

Table 1.1 Summary of currently recognizable DNA transposons in the human genome with copy number > 100

Superfamily	Families	No. of families	Copy No.
hAT	Autonomous		
	Blackjack, Charlie, Cheshire, Zaphod	19	46,133
	Nonautonomous		
	Arthur1, FordPrefect, MER102, MER106, MER107, MER112, MER113, MER115, MER117, MER119, MER1, MER20, MER3, MER30, MER33, MER45, MER58, MER5, MER63, MER69, MER81, MER91, MER94, MER96, MER99, ORSL	52	218,059
	Total	71	264,192
MuDR	Nonautonomous		
	Ricksha	3	985
	Total	3	985
piggyBac	Autonomous		
	Looper	1	521
	Nonautonomous		
	MER75, MER85	3	1,569
	Total	4	2,090
Tc1/*mariner*	Autonomous		
	HSMAR, Tigger, Kanga	22	53,320
	Nonautonomous		
	MADE, MARNA, MER104, MER2, MER44, MER46, MER53, MER6, MER8, MER82, MER97	23	54,718
	Total	45	108,038
Unknown	MER103, MER105	2	7,567
	Total	2	7,567
	Grand Total	**125**	**382,872**

1.2.2 Analysis of DNA transposons nested into other elements

In order to obtain a first assessment of human DNA transposon families that were active during the primate radiation, we took advantage of the fine-scale evolutionary histories of L1 and *Alu* elements in the primate lineage produced by others. We used these primate-specific families as historical markers for dating DNA transposons. We reasoned that any DNA transposon inserted or nested within a primate-specific L1 element (Khan et al. 2006) or a primate-specific dimeric *Alu* element (Kapitonov and Jurka 1996) should itself be primate-specific. Using a Perl script and data from the UCSC Genome Browser RepeatMasker tables, we conducted an exhaustive search of human L1 and *Alu* retroposons that had suffered a nested insertion of a DNA transposon. A DNA transposon was considered to be nested within

one of these retroposons if: (a) the upstream and downstream retroposon fragments were within 50 bp of the 5' and 3' ends of the DNA transposon, (b) the orientation of the upstream and downstream retroposon fragments were the same, and (c) the 3' end of the upstream retroposon fragment and the 5' end of the downstream retroposon fragment were within 20 bp of each other according to matching positions in their consensus nucleotide sequence.

This analysis revealed the presence of elements representing ten distinct DNA element families that were inserted into primate-specific L1s (Table 1.2). Each of these insertions were validated by visual inspection of expected target site duplications (TSDs) flanking the DNA transposon insertion based on alignment to the family consensus sequence (one example per family is shown in Table 1.2). The only exception was MER107 elements, whose insertion into L1 created a short deletion (from 11 to 19 bp) at the integration site accompanied by the co-insertion of unrelated 'filler' DNA (11 to 27 bp), instead of the 8-bp TSD canonical of hAT elements (Figure 1.1). The youngest L1 elements that suffered a DNA transposon insertion belong to the L1PA8A family, estimated to be between 42 and 50 My old (Khan et al. 2006). L1PA8A elements suffered insertions from two separate MADE1 elements previously proposed to be among the youngest DNA element families (Lander et al. 2001) (Table 1.2).

Table 1.2 DNA element insertions into primate-specific L1s, *Alu*s and LTRs

DNA element	Superfamily	LINE	Target site duplication	Age of LINE	Avg % div. of LINE	Position (May 2004 Assembly)	No. of nestings found
MER85	*piggyBac*	L1PA10	TTAA/TTAA	46.4[a]	13.03	chr11:87694072–87694198	4
MER107	hAT	L1PB3	None	73.5[a]	17.84	chr2:166664067–166664265	1
MER75B	*piggyBac*	L1MA2	TcAA/TtAA	65.8[a]	15.75	chr7:144252811–144253077	1
MADE1	Tc1/*mariner*	L1PA8A	TA/TT	41.7[a]	12.92	chr6:91273479–91273559	2
HSMAR1	Tc1/*mariner*	L1MA1	TA/TA	61.6[a]	15.58	chrX:85338493–85339772	2
Charlie3	hAT	L1PA16	CTgTATCC/CTaTATCC	79.7[b]	18.35	chr12:83988086–83990697	1
MER30	hAT	L1MA1	TTCTAATG/TTCTAATG	61.6[a]	15.58	chr11:43680827–43681054	2
MER30B	hAT	L1MA3	TCCaGGAT/TCCtGGAT	68.1[a]	15.96	chr3:117825182–117825366	1
MER75	*piggyBac*	L1PA13	TTAA/TTAA	65.8[a]	15.75	chr6:110019741–110020255	3
MER1B	hAT	L1MA1	GTTTAGaT/GTTTAGcT	61.6[a]	15.58	chrX:33568269–33568602	2

DNA element	Superfamily	*Alu*	Target site duplication	Age of *Alu*	Avg. % div. of *Alu*	Position (May 2004 Assembly)	No. of nestings found
MER85	*piggyBac*	*Alu*Jb	TTAA/TTAA	56[c]	16.57	chr1:15696507–15696630	1
MER107	hAT	*Alu*Jo	None	60[c]	17.43	chr16:15634659–15634717	1
MER75B	*piggyBac*	*Alu*Jb	TTAA/TTAA	56[c]	16.57	chr1:202795190–202795404	1
MADE1	Tc1/*mariner*	*Alu*Sx	TA/TA	39.8[d]	12.17	chr3:187935090–187935161	4
MER30	hAT	*Alu*Jo	GGCTAGAG/GGCTAGAG	60[c]	17.47	chr8:25981652–25981851	6
MER1A	hAT	*Alu*Jb	GCTGGGAc/GCTGGGAt	56[c]	16.57	chrX:5362201–5362641	1
MER1B	hAT	*Alu*Jb	GCTTAAaC/GCTTAAgC	56[c]	16.57	chr19:35538374–35538704	1

DNA element	Superfamily	LTR	Target site duplication	Avg. % div. of LTR	Position (May 2004 Assembly)	No. of nestings found
MER75B	*piggyBac*	MSTB	TTcA/TTAA	16.86	chr6:520183–520422	1
MADE1	Tc1/*mariner*	THE1B	TA/TA	14.75	chr21:27628868–27628948	1
MER30	hAT	MSTA	TGCTACAC/TGCTACAC	14.75	chr9:117723954–117724179	2
MER75	*piggyBac*	MSTA	TTAA/TTAA	14.75	chr5:36127535–36128098	2
MER1A	hAT	MSTA	GCTAAACC/GCTAAACC	14.75	chr5:35215654–35216182	6
MER1B	hAT	MSTA	GGTTTAGT/GGTTTAGT	14.75	chr7:35906363–35906685	6

[a]Age from Khan et al. 2006.
[b]Age from Smit et al. 1995 and Khan et al. 2006.
[c]Age from Price et al. 2004.
[d]Age from Xing et al. 2004.

Figure 1.1 MER107 insertions into L1 and *Alu*s provoked a short deletion accompanied by the co-insertion of filler DNA. For each alignment, the first line shows the flanking sequences of a MER107 element nested into a L1 or *Alu* element in black, while the filler DNA accompanying the insertion at one of the junction is shown in red. The boundaries of the MER107 were determined according to the RepeatMasker output and confirmed by manual alignment to the consensus sequence. For the nested insertions into L1, the bottom line shows the sequence of a related L1 copy from the same subfamily located elsewhere in the genome. For the nested insertion into *AluJo*, the bottom line shows the consensus *AluJo* sequence. The sequence presumably deleted upon MER107 is in blue. Hyphens denote gaps.

In each case, we found that the length of the nested DNA transposons was similar to the length of its respective consensus sequence and that the nucleotide divergence of the L1 copy to its consensus sequence was congruent with the age of the corresponding L1 subfamily. For example, four distinct MER85 elements were found inserted into L1PA10 elements (one example is reported in Table 1.2). A TSD of the tetranucleotide sequence TTAA, a hallmark of piggyBac transposons, could be associated with each insertion. The length of the nested MER85 elements ranged from 126 bp to 130 bp, which is in good agreement with the length of the family consensus sequence (140 bp). The nested MER85 copies differ by 6.3% to 9.3% from their consensus sequence, while the average divergence for the family is 7.3% (as calculated from the May 2004 Repeatmasker files from the UCSC Genome Browser, see Methods). The L1PA10 elements that suffered insertions by MER85 elements were 8.5% to 23.9% divergent from their consensus (with three of the four being 15.3% diverged or less), which is consistent with the average 14.7% pairwise divergence of the L1PA10 subfamily (Khan

et al. 2006). These data indicate that both nested DNA transposon and disrupted L1 elements are fully representative of their respective families.

We found that 6 out of the ten DNA element families that had copies inserted into primate-specific L1 elements also comprise copies nested into dimeric *Alu* elements (Table 1.2), all of which are known to be primate-specific (Kapitonov and Jurka 1996). One family, MER1A, was found to be nested within a dimeric *Alu* (*AluJb*) but was not found to be nested within a primate-specific L1. Here again, each DNA transposon insertion was validated as a bonafide transposition event by the presence of expected TSD (Table 1.2), except for MER107, which once again created a short deletion upon insertion rather than a typical TSD (Figure 1.1). The youngest *Alu* element that suffered a DNA element insertion was a member of the *AluSx* subfamily, which amplified ~44 Mya (Xing et al. 2004). Four unambiguous instances were found of an *AluSx* element that had suffered an insertion of a MADE1 transposon (one example is shown in Table 1.2). Each MADE1 was integrated at a different position within each *AluSx* element, indicating that they all resulted from independent MADE1 transposition events rather than from propagation of a composite element.

Though the evolutionary history of primate LTR elements has not been analyzed as fully as those of L1 and *Alu* elements, we drew upon the available literature to provide further validation for the nested insertion analysis. Smit demonstrated that 5 families of LTR retrotransposons (THE1A, THE1B, THE1C, MSTA and MSTB) were primate specific (Smit 1993). We found that 6 of the 11 DNA elements that had nested within the primate-specific L1 and *Alu*s had nested within these LTR elements (Table 1.2). Each insertion was validated by the presence of the expected TSD. The youngest primate-specific LTR to suffer an insertion from a DNA element was THE1B.

1.2.3 Cross-Species Genomic Analysis of Orthologous Insertions

To refine the age of the DNA transposon families, we turned to another method that did not rely on sequence divergence. It is possible to ascertain when individual TE insertions occurred by investigating their presence/absence at orthologous genomic regions in multiple

species whose phylogenetic relationships are confidently established. If an element is present and fixed within the population in one species, but is absent at the orthologous position in another species, then the element must have been transposing some time after the split of the two species. Note that for DNA transposons, which transpose through a cut-and-paste mechanism, this pattern could be caused by either the insertion of the element or by its excision in one of the two species examined. In either case, nonetheless, the absence of the element in one of the two species can be interpreted as a manifestation of transposon activity after the divergence time of the two species. Conversely, if an element is present at orthologous positions in two different species, it is almost certain that this insertion predates the divergence of the two species. This is because the probability of two elements of the same family inserting at the exact same position independently in two different lineages is incredibly small, especially for DNA elements, as these occur in relatively low copy number families in mammalian genomes.

Taking advantage of the ongoing NISC-NIH Encode Project and of other genomic resources accessible through the UCSC Genome Browser, we assessed the presence/absence of 794 human DNA transposon copies at orthologous positions in at least 8 (and up to 10) other mammalian species, including 3 primate species (see Methods). These elements represent 111 families out of 125 families known in the human genome.

We found members of 11 DNA element families that were present at orthologous positions in human, Rhesus macaque (Macaca mulata) and marmoset (Callithrix jacchus), but absent in the galago (Otolemur garnettii), a prosimian primate. We were able to identify 'empty sites' in the respective orthologous galago loci for members of each DNA element family (one example per family is shown in Table 1.3) except for MER75B since there is only one copy of MER75B within an Encode region and there is a large deletion within the galago lineage at that locus. These were clear 'empty sites', with only one copy of the TSD and no additional sequences indicative of transposon excision. DNA transposon excisions are typically imprecise in that they leave behind one of the two TSD and/or a few terminal nucleotides of the

transposon at the excision site (Plasterk 1991; Craig et al. 2002). This data suggests that these elements were inserted in the anthropoid lineage, rather than excised in the galago lineage. We assume that these individual copies are representative of their respective families because their individual percent divergences are all within the 95% confidence interval for average divergence of the family (average divergence plus or minus 1.96 standard deviations). As such, these elements do not represent statistical outliers. These data demonstrate that at least 11 families of human DNA transposons were transpositionally active after the split of prosimians and anthropoid primates, or during the last ~63 My (Goodman 1999). Seven of these 11 families also include copies nested within L1 or *Alu* elements known to be primate-specific (see above and Table 1.2). Assuming that these individual copies are representative of their respective families and that their activity is contemporary to the activity of their entire family, these 11 families make up a total of 23,570 transposons in our genome. Therefore, these data imply that many thousands of DNA transposons were inserted in the lineage of anthropoid primates, that is within the last ~63 My.

Table 1.3 Pre-integration empty sites in galago

DNA element	Position (July 2003 Assembly)	Galago accession	Empty site
MER85	chr11:131136027–131136354	AC149045	gctgattaaaaccatttatg...aaatgggttaagccat gttgattag----------gccac
MER107	chr5:56187171–56187506	AC146960	ggctttgcaggggtatccaagg...agctagacaagcttgctaatt gactt----------taatg
MADE1	chr21:32755726–32756005	AC146494	atttgtaggttggtgcaaaaa...caccaacctaataaaatg cttggta----------gaatg
HSMAR1	chr18:24065341–24066812	AC146672	tatta-------tattaggttggtgca...tgcaccaacctaatatgttt cattaaaacttt----------tattttt
Charlie3	chr21:39392024–39392662	AC146737	taatacataaatacaggggtctc...ttggggaccactgcataaatacacct taatatttaaata----------tgca
MER30	chr21:34097462–34097675	AC148495	ataagcatttaaaccaagcttgtc...agattgtactgatttagacactc acaaccatttaga----------tgatc
MER30B	chr7:116791959–116792151	AC12354	tttgactgcagaaaatggatgtcccc...ttggacacagctgctctagatatttt tttga----------ctatagattattga
MER75	chr21:32810522–32810858	AC146494	cag------attaaacccttctcccatt...aaacgaaagggttaaaacac ctggaatagat----------ttaaaacac
MER1A	chr2:235151057–235151743	AC148542	gtcttagagcaggggtgccgg...tgactgctgttttagagccaga gtcctagag----------ctagg
MER1B	chr7:116073962–116074318	AC146879	aatctatctacagcagcagtccc...ggggacccctgatctacagaattg agtctacct--ag----------aattg

Members of the remaining 100 human DNA transposon families represented in the Encode regions were found at orthologous positions in human, marmoset and galago. To investigate which of these families was primate-specific, each of the 316 copies, along with at least 100-bp flanking both the 5' and 3' ends, was visually inspected in the UCSC Genome

Browser for their presence/absence at orthologous positions in at least five non-primate mammals using the Encode Comparative Genomics tracks. Phylogenetically, these five species form two separate outgroups to the primates, with the mouse/rat/rabbit lineage being closer to the primates than the cow/dog lineage (Murphy et al. 2004). Hence a transposon present in primates but absent in the 5 other mammals most likely resulted from an insertion in the primate lineage. A less parsimonious explanation for this pattern would be that the element inserted in a eutherian (i.e. placental) ancestor and subsequently excised twice independently, in the mouse/rat/rabbit lineage and in the dog-cow lineage. We found 163 elements from 23 families that were clearly present in humans as well as 3 of the 5 other primate genomes (chimp, baboon, rhesus, marmoset, and galago), but absent in all other mammals examined. Thus these 23 families were classified as primate-specific.

The remaining 77 families were represented by elements present at orthologous sites in all the primates and at least one of the other eutherian mammals; thus these families were classified as eutherian-wide. It should be noted, however, that several families (such as Charlie1 and Charlie1a) included some copies that were apparently primate-specific as well as copies present at orthologous positions in primates and at least one of the non-primate mammals. It could well be that the activity of these families initiated prior to the emergence of the primates, but continued in a primate ancestor, generating primate-specific insertions. Alternatively, lineage-specific sorting of ancient alleles with or without the insertions could also account for these patterns. In the absence of further evidence to distinguish between these possibilities, we adopted a conservative classification of such families as eutherian-wide.

To determine if the other 14 DNA element families not represented within the Encode regions were primate-specific, we searched the finished mouse, rat and dog genomes for the presence or absence of TEs at orthologous positions to the human DNA elements (see Methods). We found that 6 families of elements (MER6C, Ricksha, Ricksha_b, Ricksha_c, Tigger5a and Tigger5b) were present in primates but clearly absent in the mouse, rat and dog. These families were additionally classified as primate-specific. The other 8 families were found

to be present in the primates and in the mouse, rat or dog genomes. These families were classified as eutherian-wide. Together, the cross-species analysis of orthologous insertions suggests that a minimum of 40 distinct DNA transposon families, which accounts for 98,300 DNA elements currently fixed in the human genome, were active in the primate lineage, i.e. within the last ~80 My.

1.2.4 Age of DNA transposons based on sequence divergence

The results above allowed us to distinguish among DNA transposon families that have been active in an anthropoid (40-63 Mya), primate (63-80 Mya), or eutherian ancestor (80-150 Mya) (Goodman 1999; Murphy et al. 2004). In order to obtain a better resolution of the age of each family, we used the median of these three evolutionary periods (51.5, 72, and 115.5 My, respectively) to determine the median substitution rate per million years of all DNA transposons within each age class (anthropoid-, primate- or eutherian-specific). We also calculated a low and high substitution rate using the upper and lower bounds of the age class. This was done by calculating the corrected number of nucleotide substitutions per site in all individual copies to their respective family consensus sequence (as given by Repeatmasker), and excluding consensus CpG positions using the REV model (Tavare 1986) (see Methods). The REV, or general reversible model, calculates 5 different rate parameters based upon nucleotide composition, thereby correcting for differing nucleotide compositions between DNA element families. The same analysis was performed separately with the L1s and *Alu*s falling into the anthropoid- or primate-specific class ages (plus an additional class of less than 40 My) and four eutherian-wide L1 elements (L1PA17, PB4, and MA4-5), following the most recently published dating of L1 (Khan et al. 2006) and *Alu* subfamilies (Price et al. 2004; Xing et al. 2004).

This approach allowed us to calculate a substitution rate per million years that takes into account possible fluctuations in evolution rates during the different time periods and among the three types of elements (see Methods). The results reveal that DNA transposons, L1s and *Alu*s each have different average substitution rates in the different evolutionary periods (Table 1.4). There are statistically significant differences among the three types of TEs within the

same period. For example, within the anthropoid-specific lineage (41-63 myr), which was the only period for which mutation rates could be estimated for all three types of TEs, the differences were statistically significant ($p < 0.05$, ANOVA). These differences may be attributed to several factors, such as biased genomic distributions (for example, *Alu*s preferentially accumulate in GC-rich regions), compositional biases, replication mechanisms (reverse transcriptase for RNA elements, DNA polymerase for DNA transposons) and amplification dynamics (subfamily structure). Regardless, this data provides a rationale for treating DNA transposons separately from the other types of TEs in the human genome in these calculations, as opposed to using substitution rates estimated for other type of elements or for other neutrally evolving sequences such as pseudogenes (Robertson and Martos 1997).

Table 1.4 Substitution rates (μ) per million years

	Substitution Rates			
Group	**DNA**	***Alu***	**L1**	**P**
< 40 myr		0.3165%	0.1656%	$p > 0.05$
41 - 63 myr	0.1774%	0.3102%	0.1880%	$p < 0.05$
64 - 80 myr	0.2125%		0.2574%	$p < 0.05$
> 80 myr	0.2509%		0.1601%	$p < 0.05$

We next applied the substitution rate of each type of element for each period to estimate the age of individual families based on their average nucleotide divergence to the respective consensus sequence, excluding CpG sites (see Methods). This method differed slightly from other datings of TEs in which a single, constant substitution rate for the entire span of primate evolution was used (Price et al. 2004; Khan et al. 2006). Our estimates for the age of L1 and *Alu* subfamilies using these period specific substitution rates differed somewhat from recently published datings (Price et al. 2004; Xing et al. 2004; Khan et al. 2006) (Figure 1.2). This is to be expected since all three published datings used the Kimura 2-parameter correction not the REV model (Kimura 1980). Overall, L1 elements appeared to be slightly older, due to their higher consensus AT content, and *Alu* elements appeared to be slightly younger, due to

their lower consensus AT content, using the REV model than using the Kimura model. The resulting estimated ages of human DNA transposon families less than or equal to 80 My is given in Table 1.5 and the dating for all 125 families is available in Appendix A. Table 1.5 gives the median age for the family and the upper and lower boundaries of the age.

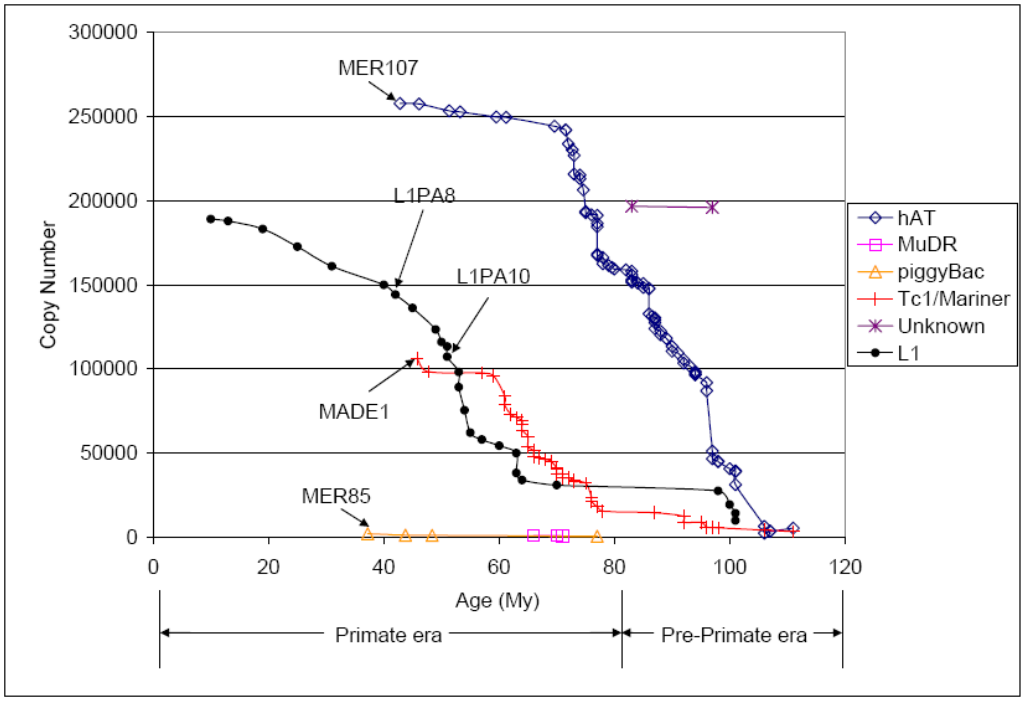


Figure 1.2 Cumulative copy numbers for DNA element superfamilies and L1 elements according to age. The age of each family is plotted against the cumulative copy number of the superfamily. See Results and Methods sections for details on the calculation of the age for each family. (A) hAT superfamily elements were more intensively active during the pre-prosimian era (> 63 My), decreasing in proliferation during the subsequent primate radiation (< 63 My). (B) MuDR and piggyBac (e.g., MER85) superfamily elements were strictly active during the primate radiation. (C) Tc1/*mariner* (e.g., MADE1, Tigger1) superfamily elements were active before the primate radiation, but experienced a marked burst of activity during the primate radiation.

Table 1.5 DNA elements less than 80 million years

Categories	
1	40 - 63 My - Anthropoid-Specific
2	64 - 80 My - Primate-Specific
3	80 - 150 My - Eutherian-Specific

Category	RepName	Corrected Age (My) (Low Age - High Age)	% Div (excluding CpGs)	AT Content of Consensus Seq (%) (excluding CpGs)
1	MER85	37 (29-46)	4.6	64
1	MER107	43 (33-52)	6.0	59
1	MER75B	44 (34-54)	6.6	66
1	MADE1	46 (36-56)	7.5	66
1	MER30	46 (36-56)	7.0	64
1	HSMAR1	48 (37-58)	7.9	64
1	MER75	48 (38-59)	8.5	68
1	Charlie3	51 (40-63)	8.7	59
1	MER1A	53 (41-65)	9.3	56
2	HSMAR2	57 (48-60)	14.8	64
2	Tigger1	59 (50-62)	14.8	64
1	MER30B	60 (46-73)	7.7	62
2	Tigger3b	61 (51-64)	10.9	67
2	Tigger2	61 (52-64)	16.7	63
1	MER1B	61 (48-75)	11.8	51
2	Tigger4(Zombi)	62 (52-65)	10.0	64
2	MER46B	63 (53-67)	17.0	66
2	MER44A	64 (54-67)	13.0	66
2	Tigger2a	64 (54-67)	16.7	63
2	Tigger3(Golem)	64 (54-67)	10.7	66
3	MER53	65 (46-85)	21.2	70
2	Tigger5a	65 (55-69)	16.1	65
2	Ricksha_c	66 (55-69)	18.8	63
2	MER46A	66 (55-69)	15.0	64
2	MER6	66 (56-70)	14.5	63
2	MER6A	67 (56-70)	15.2	62
2	MER44C	68 (57-72)	15.0	62
2	Tigger7	69 (58-73)	12.1	62
3	Tigger6b	70 (49-91)	21.7	66
2	Ricksha_b	70 (59-74)	29.7	64
2	Tigger5	70 (59-74)	19.0	64
3	MER96B	70 (49-91)	20.1	71
2	MER44D	70 (59-74)	18.0	62
2	Ricksha	71 (59-74)	13.4	63
2	MER6C	71 (60-75)	22.2	66
2	MER2B	71 (60-75)	18.5	62
3	Charlie5	72 (50-93)	23.2	76

Table 1.5 - *Continued*

2	MER44B	72 (60-75)	17.0	64
3	MER58D	72 (51-94)	26.1	69
3	MER33	72 (51-94)	25.0	74
3	Tigger6	73 (51-94)	25.9	68
3	MER45	73 (51-95)	21.0	59
2	Tigger5b	73 (61-77)	18.2	61
3	MER3	73 (51-95)	24.5	68
3	Tigger6a	73 (51-95)	26.7	66
3	MER45A	73 (52-95)	21.2	59
3	MER58	74 (52-96)	24.2	67
3	MER58B	74 (52-96)	18.3	65
2	MER2	75 (63-79)	22.9	63
3	MER45B	75 (52-97)	26.1	65
3	Cheshire	75 (53-97)	30.3	69
3	MER58A	75 (53-98)	23.1	62
2	MER8	76 (64-80)	17.2	57
3	MER45R	76 (53-99)	26.6	69
2	MER6B	76 (64-80)	18.9	65
2	MER82	76 (64-80)	19.0	64
3	Charlie1	77 (54-99)	29.3	69
3	Looper	77 (54-99)	31.3	70
3	ORSL	77 (54-100)	25.9	71
3	Charlie1b	77 (54-100)	30.9	68
3	MADE2	77 (54-100)	23.3	72
3	MER20	77 (54-100)	15.5	53
3	MER63B	77 (54-100)	26.4	69
2	Tigger5c	78 (65-82)	21.6	60
3	Charlie1a	78 (55-101)	29.1	67
3	MER63D	78 (55-101)	26.6	71
3	MER119	79 (55-102)	29.2	64
3	MER63C	80 (56-104)	29.7	69
3	MER106B	80 (56-105)	29.4	64
3	MER106A	82 (58-107)	21.9	60
3	MER96	83 (58-107)	23.9	48
3	MER105	83 (58-107)	20.3	62
3	MER99	83 (58-107)	24.0	67
3	MER58C	83 (58-107)	25.9	65
3	MER63	83 (58-108)	30.0	65
3	MER97a	84 (59-109)	19.7	66
3	MER97b	84 (59-109)	32.1	71
3	Charlie4a	85 (60-110)	33.0	70
3	MER91	85 (60-111)	38.3	55
3	MER5A1	86 (60-111)	20.4	56
3	MER5C	86 (60-112)	33.5	73
3	Charlie6	86 (61-112)	34.2	66
3	Charlie4	87 (61-112)	33.3	68
3	MER97c	87 (61-112)	24.7	71

Table 1.5 - *Continued*

3	Charlie9	87 (61-113)	39.8	67
3	MER104	87 (61-113)	26.6	71
3	MER81	87 (61-113)	25.5	47
3	Zaphod2	87 (61-114)	20.7	69
3	Charlie10	88 (62-114)	35.8	68
3	Zaphod	88 (62-114)	34.3	70
3	Arthur1	89 (63-116)	37.4	66
3	MER69A	90 (63-117)	31.8	60
3	MER63A	90 (63-117)	43.6	58
3	Kanga1	92 (64-119)	39.2	68
3	Kanga1c	92 (64-119)	30.0	66
3	MER112	92 (64-119)	39.8	69
3	Charlie2a	92 (65-120)	39.1	70
3	BLACKJACK	92 (65-120)	38.9	69
3	Charlie2	93 (65-121)	35.0	68
3	FordPrefect	94 (66-122)	42.4	53
3	MER45C	94 (66-122)	20.5	45
3	MER69B	94 (66-122)	34.9	65
3	MER94	94 (66-123)	38.5	58
3	MER46C	95 (66-123)	36.9	65
3	MER5A	96 (67-125)	77.3	52
3	Kanga1a	96 (67-125)	42.0	65
3	Charlie7	96 (68-125)	27.4	70
3	MER91B	97 (68-126)	40.9	59
3	Kanga1b	97 (68-126)	33.1	64
3	MER103	97 (68-126)	39.4	66
3	MER113	97 (68-126)	36.9	70
3	Kanga2_a	98 (69-127)	41.9	68
3	MER20B	98 (69-128)	42.7	63
3	FordPrefect_a	98 (69-128)	57.5	45
3	MER91C	100 (70-130)	25.6	60
3	Charlie2b	101 (71-131)	45.7	63
3	Charlie8	101 (71-131)	40.5	61
3	MER5B	101 (71-132)	53.0	53
3	Tigger8	106 (74-137)	29.1	68
3	MER115	106 (74-138)	42.3	46
3	MER102b	106 (75-138)	65.4	58
3	MER91A	107 (75-139)	45.8	44
3	MER102a	107 (75-139)	72.6	56
3	MER117	111 (78-144)	33.2	53
3	MARNA	111 (78-145)	46.2	62

This data was used to generate a plot of the age of DNA transposons and L1s as a function of their copy number (Figure 1.2). This representation shows that our dating of DNA

transposons is in good agreement with the nested insertion analysis, but provides a better resolution of the age of individual family. For example, MER85 and MADE1 are among the youngest DNA elements, with estimated ages of ~37 and ~46 Mya, respectively, and both families included members nested within L1PA10 elements (Table 1.2), a subfamily that we dated as ~43 My old (Khan et al. 2006). Conversely, there are no DNA transposon families significantly younger than L1PA8, as indeed we found no instances of any DNA element nested within L1PA8 or younger. But we could detect several L1PA8 elements nested within MER85 and MADE1 transposons (data not shown).

Our dating of individual DNA transposon families based on sequence divergence and calculated age is also largely congruent with the cross-species analysis. All elements classified as eutherian-wide from this analysis were found to be older than 65 My based on sequence divergence, with all but one family (MER53) being older than 70 My. MER53 is an outlier due to its unusually high 70% AT content (Table 1.5). All DNA transposon families classified as primate-specific by the cross-species analysis were estimated to be between 57 and 78 My based on sequence divergence.

Figure 1.2 reveals that there were two bursts of DNA transposon activity in the time period between the mammalian radiation and the split of New World Monkeys from the primate ancestor. The first peak is the most pronounced and involves primarily members of the hAT superfamily and spanned a period of ~40 My from a pre-primate era to early primate evolution (~70 Mya). The second subsequent peak is strictly primate-specific (from 80 to 63 Mya) and mostly implicated Tc1/*mariner* elements, although the greatest diversity of DNA transposons was active during this time, including hAT, MuDR and PiggyBac elements (Figure 1.2). The relatively more recent activity of Tc1/*mariner* elements compared to the most abundant hAT elements has been previously noticed (Lander et al. 2001). Our data indicates that the burst of Tc1/*mariner* activity reached a plateau which seems to coincide with the emergence of anthropoid primates, at ~63 Mya. Three superfamilies (hAT, Tc1/*mariner*, PiggyBac) continued

to be active in the anthropoid lineage, but there seems to be a sudden loss of activity of all DNA transposons shortly after the emergence of the new world monkeys ~40 Mya.

1.3 Discussion

While the evolutionary history of human *Alu* and L1 retrotransposons has been studied intensively, the history of DNA transposons has largely been overlooked. In this study we have combined three different approaches to determine the average age of all 125 DNA transposon families known in the human genome. The results of the three approaches converge to reveal that a substantial fraction of human DNA transposon families (at least 39 and up to 69 families, see Tables 1.5 and 1.6), representing at least ~98,000 elements in our genome, were transpositionally active in the primate lineage. Below we first discuss the value of combining various methods for estimating the age of TEs, then turn to the specific implications of our findings for primate genome evolution and for understanding the forces underlying the amplification dynamics of TEs in mammalian genomes.

1.3.1 Combined methods provide a detailed estimate of the age of human DNA transposons

As new genome sequences are released, different methods for dating transposable elements are being developed that allow greater accuracy in estimating the age of TEs (Price et al. 2004; Salem et al. 2005; Caspi and Pachter 2006). This is a crucial aspect of genome research because TEs not only provide a rich fossil record to determine the pace and mode of molecular evolution (Waterston et al. 2002), but they are also major players in the structural evolution and regulation of genes and genomes (Feschotte et al. 2002; Deininger et al. 2003; Kazazian 2004; Cordaux et al. 2006b). To our knowledge, our study is the first that use a combination of three independent methods to evaluate the age of a broad population of TEs on a genome-wide scale.

The first method, nested insertion analysis, capitalizes on the well-characterized history of L1 and *Alu* elements and provides an estimation of the relative age of TE families. This method does not rely on the molecular clock and can be performed even in the absence of

genomic sequences for other closely related species. The shortcoming of this approach is that not every TE family member will necessarily suffer an insertion from every other TE family, thereby leading to gaps in the data, especially for TE families with low copy numbers. The second method (cross-species analysis of orthologous insertions) takes advantage of the large amount of sequence data recently generated for several mammalian species and is also independent of the molecular clock. These first two methods deliver a rough, yet robust evaluation of the time periods when the elements were active. This, in turn, provided us the means to calibrate the molecular clock at three different evolutionary time points, which allowed the refinement of the age of each human DNA transposon family based on nucleotide divergence of individual copies to their ancestral consensus sequence. There is only partial overlap between the results gathered by the three methods (Figure 1.3), which emphasizes the value of combining the three methods to derive a reliable history of the entire population of human DNA transposons.

Figure 1.3 Comparison of three independent methods for dating DNA transposons. Eleven distinct families of DNA transposons were shown to be primate-specific by all three methods. Twenty-nine additional DNA element families were found to be primate-specific by both the percent divergence and cross-species genomic analysis methods. Twenty-nine families predicted to be primate-specific based solely on percent divergence and age were determined to be eutherian-wide by cross-species analysis.

Of the 125 DNA element families currently recognizable in the human genome, a total of 11 families could be classified as primate-specific by all three methods (Figure 1.3). Since nested insertion analysis does not allow us to examine all DNA transposon families, but only a subset of them, one cannot expect a complete overlap between the results produced by the three different methods. Sixty-nine families of DNA transposons were predicted to be primate-specific based on sequence divergence data and thus, calculated age, alone (Table 1.5) and 40 of these families were confirmed to be primate-specific by at least one of the two alternative methods. Hence, we believe that this set of 40 families provides a reliable, yet conservative estimate of DNA transposons that were integrated during primate evolution. The corresponding families range in age from MER85 (37 My) to Tigger5c (78 My) and have contributed 98,300 elements (totaling ~33 Mb of sequence) to the human genome (Figure 1.4). Furthermore, the results of divergence and cross-species and/or nested insertion analysis confirm that nearly one-fourth of these transposons (23,462 elements, ~5 Mb) have likely been inserted since the split of anthropoid primates from prosimian primates, or within the last ~63 My (Goodman 1999).

Thirty families were predicted to be primate-specific based on their sequence divergence and calculated age but were shown to be eutherian-wide by cross-species analysis. However, for some of these families, such as MER53, Charlie5 and MER33 (65, 72, 72 My, respectively), we could detect copies present at orthologous positions in all eutherian species examined, which strongly indicates that at least a subset of family members inserted prior to the divergence of placental mammals. Thus, sequence divergence alone may not always be an accurate measure of the age of TEs. It could be that, for various reasons, members of these families evolve more slowly overall than other families. A non-mutually exclusive explanation is that these families include a subfamily of primate-specific elements as well as older elements. Further analyses are required to distinguish between these possibilities.

Table 1.6 Comparison of three methods for dating DNA TEs

Method	Number of Primate Specific DNA Tes	Percent Divergence of Oldest Primate Specific Family	Oldest Primate Specific DNA TE
Average Percent Divergence	69	29.4%	MER106B
Analysis of Nested Insertions	11	11.8%	MER1B
Cross-Species Genomic Analysis of Orthologous Insertions	40	21.6%	Tigger5c

1.3.2 A general extinction of DNA transposon activity in the anthropoid lineage

An interesting phenomenon is observed when looking at the overall history of DNA transposons in the mammalian and primate lineages (Figure 1.4). Eighty-five families, or ~291,000 DNA transposons, are shared between primates and other mammals. In contrast, 29 families, or ~74,000 elements, were active specifically in primates prior to the split of emergence of anthropoids, and 11 families, or ~23,000 elements, were integrated in anthropoid species. Thus there was a steady decline in the activity of DNA transposons during primate evolution (see Figures 1.2 and 1.4). According to our combined age calculations, we found no evidence for DNA transposon families significantly younger than the divergence of new world monkeys, that is ~40 Mya (Figure 1.2 and Table 1.5). Furthermore, we conducted a systematic survey for the presence/absence of human DNA transposons at orthologous positions in the nearly complete genome of the Rhesus macaque (an old world monkey) and could not uncover a single conclusive instance of a DNA transposon copy present in human, but missing in the macaque (data not shown). Thus there is no evidence for the activity of any DNA transposons after the emergence of old world monkeys. The last active DNA transposon families represented in the human genome seem to have all became extinct in the relatively short evolutionary window (~23 My) that separated prosimians and new world monkeys (Figure 1.2). Yet, our study predicts that at the dawn of this extinction, some ~40-55 Mya, there were at least 11 families from three different superfamilies active in the anthropoid genome (Figure 1.2, Table

1.5). The majority of these elements were from the hAT superfamily (6 families), while 2 Tc1/*mariner* and 3 piggyBac families comprised the rest of this group (Figures 1.2 and 1.4). This suggests that at least 3 distinct sources of transposases, with some of them represented by hundreds of seemingly intact copies (Robertson and Zumpano 1997; Cordaux et al. 2006b) were shutdown around the same evolutionary period.

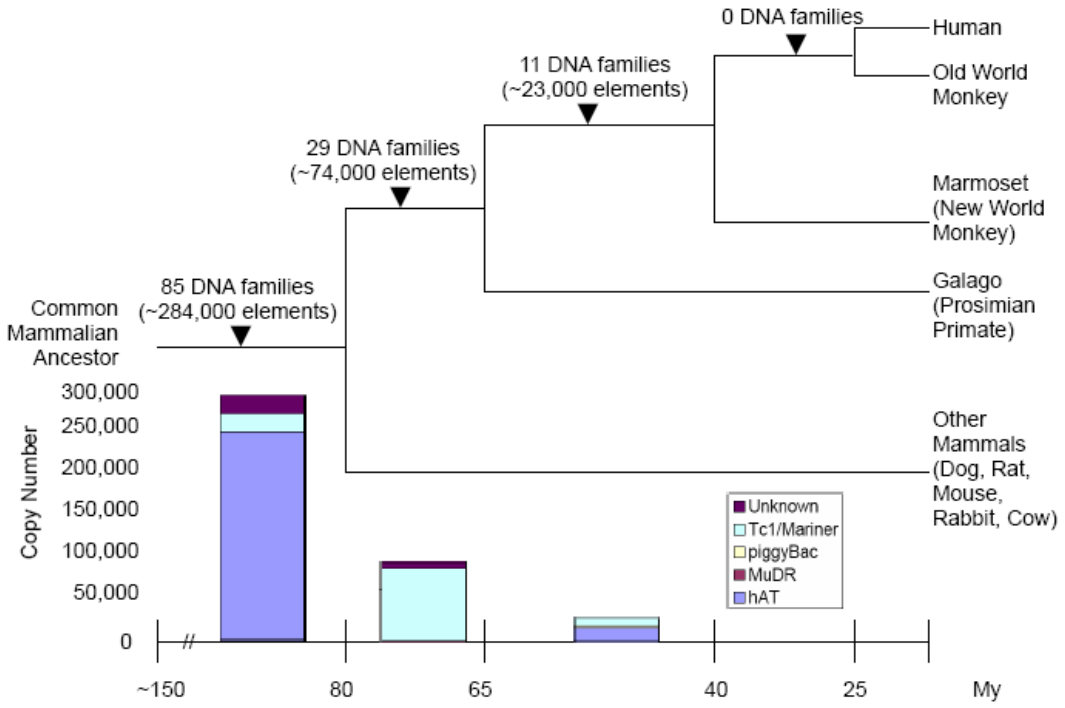


Figure 1.4 Summary of the activity of DNA transposons through primate evolution. The bar graph at the bottom of the figure represents the number of DNA elements active during each major lineage, broken down per superfamily. Note that no DNA elements were found to be active after the emergence of new world monkeys.

What could have provoked the extinction of DNA transposons that would not have affected the propagation of L1, which continued to thrive even after the emergence of new world monkeys (Khan et al. 2006) (Figure 1.2)? A distinctive feature of the life cycle of DNA transposons is their apparent propensity for horizontal transmission (Silva and Kidwell 2000; Robertson 2002). Theory predicts that horizontal transfer is indeed critical for the maintenance

of DNA transposons (Hickey 1982; Hartl et al. 1997). By contrast, horizontal transfer of LINEs occurs rarely, if ever (Eickbush and Malik 2002) and it is probably not an essential component to their maintenance (Wei et al. 2001; Kulpa and Moran 2006). It is tempting to speculate that the extinction of the DNA transposon population in the anthropoid lineage was linked to a sudden incapacity of these elements to undergo horizontal transmission. This could be due to the emergence of a host barrier aimed against the cellular entrance of TEs and other forms of invasive DNA. This would also explain the parallel regression of endogenous retroviruses during the same period of primate evolution (Lander et al. 2001). Interestingly, several defense mechanisms have been recently characterized that restrict retroviral activity in primates (Emerman 2006; Zennou and Bieniasz 2006). It could be that similar mechanisms have also compromised the propagation of DNA transposons in anthropoid primates.

1.3.3 Contribution of DNA transposons to primate genome evolution

One of the most striking findings of our study is that at least ~74,000 of the DNA transposons now fixed in the human genome (~33 Mb of DNA) were integrated during a period of less than 17 My, prior to the emergence of prosimian primates (~63 Mya) but after the divergence of a primate ancestor from the closest mammalian clades represented in our dataset (rat, mouse and rabbit; ~75-85 Mya, see Figure 1.2). This is almost twice the number of L1 elements inserted during the same period and now fixed in the human genome (~43,000 elements from the subfamilies L1PA15-16, L1PB3, L1MA2-3). Clearly, early primate evolution was a period of intense activity for DNA transposons. During the next phase of the primate radiation (63-40 Mya), i.e. after the split of prosimians, but prior to the emergence of new world monkeys, we estimated that ~23,000 DNA elements were inserted and fixed in the human genome, adding at least ~5 Mb to an ancestral anthropoid genome (Figures 1.2 and 1.4). Hence, the quantitative contribution of DNA-mediated transposition to the primate genome is far from being negligible.

Prior to this study, the history of human DNA transposons has been largely neglected relative to those of retroelements (*Alu*s and L1s). One reason for this is the common belief that

active DNA transposon families have been long extinct and that they are currently only represented by very ancient molecular 'fossils' immobilized in the genome. Indeed, unlike *Alu*s and L1s (Deininger et al. 2003), there are no known cases of de novo insertion of any human DNA transposon. Our results support the conclusion that there has been little, if any, activity of DNA transposons in the ape lineage. On the other hand, our study demonstrate that many thousands of DNA elements have integrated and become fixed during the first half of primate evolution and that several high copy number families with over 90% nucleotide identity among copies remain in the human genome (see average sequence divergence in Table 1.5). It is tempting to speculate that these primate-specific bursts of DNA transposition have had a strong impact on the structural evolution of primate genomes. DNA transposons have been frequently implicated in chromosomal rearrangements in plant and animal species, including deletions, inversions, duplications, translocations and chromosome breakage mediated by inter-element recombination or aberrant transposition events (Lim and Simmons 1994; Caceres et al. 1999; Gray 2000; Zhang and Peterson 2004). Given the medical and evolutionary importance of chromosomal rearrangements in humans (Inoue and Lupski 2002; Eichler and Sankoff 2003; Feuk et al. 2006), the possible role of DNA transposons in shaping primate genomes warrants further investigation.

1.4 Methods

1.4.1 Calculation of average percent divergence from RepeatMasker output

The average percent divergence of each transposable element family was calculated using the RepeatMasker rmsk files from the UCSC Genome Browser for the May 2004 assembly. The percent divergence (milliDiv) of each distinct element within a transposable element family was weighted by the length of the element. The average percent divergence was weighted to account for the vast difference in sizes between currently recognized elements of the same family. To calculate the average percent divergence for each family, the percent divergence calculated by RepeatMasker was multiplied by the length of the element and the

sum of all elements in each family was divided by the sum of the total length of all elements in the family.

1.4.2 Nested insertions

To find nested insertions, the RepeatMasker files from the UCSC Genome Browser and a Perl script were used to automate the analysis. Each *Alu*, L1, and DNA element in the RepeatMasker files was evaluated to see if there was a repeat within 50 bp upstream and a repeat with 50 bp downstream. If so, the upstream and downstream repeats were compared to see if the repName and Strand fields were the same. Next, the repeat positions of the ends of the upstream and downstream TEs were examined. If the upstream and downstream TEs both began within position 1 to 20 or both ended within 20 bp of the consensus length, the case was discarded, suggesting that they might represent a cluster of two elements of the same family independently inserted in close proximity and surrounding another element, rather than a single element disrupted by a nested element. Next the coordinates of the two flanking repeats were checked to verify that the end of the first repeat was within plus or minus 20 bp of the start of the second repeat according to the family consensus sequence. If all of these conditions were met, the *Alu*, L1, or DNA element was classified as nested within each other.

1.4.3 Cross-species genomic analysis of orthologous insertions

To search for the presence or absence of DNA elements in marmoset and galago, sequences for each human DNA element present in an ENCODE region (http://www.nisc.nih.gov) were retrieved from the UCSC Genome Browser in September 2005. BAC sequences for each of the orthologous Encode regions in the marmoset and galago were retrieved from Genbank (http://www.ncbi.nlm.nih.gov/Genbank/) using the accession ID given on the Encode website. From these sequences, custom BLAST databases were built for both the marmoset and galago using BLAST Version 2.2.11 (Altschul et al. 1990) and the sequences from the Genbank accessions. Each human DNA transposon, along with 100-bp of flanking sequences, were used as queries in BLASTn searches of the marmoset and galago custom databases. A repeat was classified as present at the orthologous position if at least one of the

two flanking regions and at least 50% of the element from the human sequence were found in the marmoset or galago.

1.4.4 Calculation of substitution rates and dating according to sequence divergence

The May 2004 human genome sequence (hg17) was downloaded from the UCSC Genome Browser. TEs were masked locally using RepeatMasker version 3.1.5 with the March 14, 2006 library from RepBase Update. A Perl script was used to parse the RepeatMasker align files and generate a single, concatenated sequence for each different chromosomal repeat along with the corresponding consensus sequence. The concatenated sequences had all CG dinucleotides (for + strand) and GC dinucleotides (for – strand), as well as non-ATGC characters, removed. These chromosomal repeat and consensus sequences were then combined and analyzed using PAML version 3.15 (Yang 1997). Each file was analyzed using the REV model with the clock = 1 option. The corrected number of substitutions per site was calculated as ½ of the branch length since the consensus sequence does not evolve.

To calculate the substitution rate, the corrected substitutions per site was divided by the median age for the class (anthropoid-, primate-, or eutherian-specific). The rate for each TE within the age class was weighted by the percentage of the total number of bases that TE comprised of the total base length for the entire class. These weighted rates were then summed, giving a corrected substitution rate for the entire class. The age of the family was calculated by multiplying the corrected substitution rate by the corrected percent divergence for the family.

CHAPTER 2

REPEATED HORIZONTAL TRANSFER OF A DNA TRANSPOSON IN MAMMALS AND OTHER TETRAPODS

2.1 Introduction

Lateral or horizontal transfer (HT), the transfer of genetic material between reproductively isolated species, is a frequent occurrence in prokaryotes, with selfish and mobile genetic elements such as phages, plasmids and transposons serving as the primary vehicles for HT of prokaryotic genes (Ochman et al. 2000). In contrast, reports of HT are scarce in eukaryotes and most cases of nuclear acquisition implicate transfers from prokaryotes or endosymbionts (Hotopp et al. 2007; Andersson 2005; Loftus et al. 2005; Morrison et al. 2007; Gladyshev et al. 2008). The best-documented instances of HT between the nuclear genomes of multicellular eukaryotes involve mobile genetic elements, and in particular class 2 or DNA-mediated transposons (Syvanen and Kado 2002; Silva et al. 2004). Thus far, conspicuous cases of HT of DNA transposons have been detected among insects (Silva et al. 2004; Robertson 2002; Hartl et al. 1997; Yoshiyama et al. 2001; Loreto et al. 2008), fish (de Boer et al. 2007) and, in one example, between plants (Diao et al. 2006). Germline invasions by retroviruses have been documented for several mammals (Han et al. 2007; Gifford and Tristem 2003; Yohn et al. 2005; Tarlinton et al. 2006) and there is mounting evidence supporting the horizontal introduction of a snake retroposon in ruminants (Kordis and Gubensek 1998; Piskurek and Okada 2007). However, to our knowledge, there has been no substantiated report of HT of DNA transposons in mammals. Here, we present unequivocal evidence for the repeated HT of a DNA transposon family named SPACE INVADERS in 7 tetrapod lineages, including 5 mammalian orders.

2.2 Results

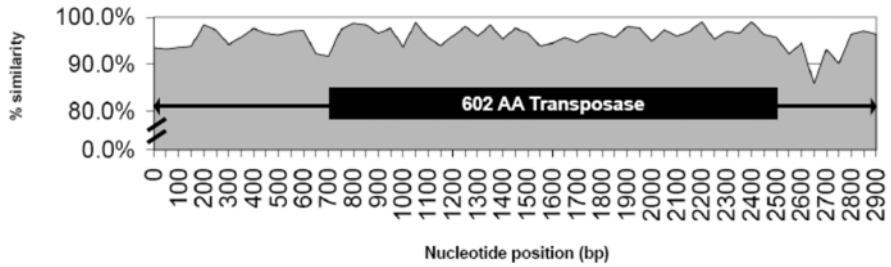
2.2.1 Discovery of SPIN transposons

While surveying DNA transposons in the draft genome assembly of the bushbaby *Otolemur garnettii*, a prosimian primate, we discovered a previously uncharacterized family of elements of the hAT (hobo/Activator/Tam3) superfamily, which we dubbed SPACE INVADERS, or *SPIN*. Alignment of 21 full-length or nearly full-length *SPIN* copies allowed us to reconstruct the putative ancestral consensus sequence for this family. The consensus sequence was 2,836-bp long and contained a single long open reading frame (ORF) encoding a 602-aa transposase with a conserved hAT dimerization domain at the C-terminus (PFAM05699). In addition, each *SPIN* copy examined was flanked by 8-bp target site duplications, another hallmark of the hAT superfamily (Feschotte and Pritham 2007).

To determine the species distribution of *SPIN*, the bushbaby consensus sequence was used as a query in Blastn searches of all GenBank databases, which currently contains whole genome shotgun sequence assembly for 33 vertebrate species, including 23 placental mammals, 1 marsupial, 1 monotreme, 3 non-mammalian tetrapods and 5 fish (see Materials and Methods). These searches yielded a large number of extremely high score hits (score > 1800, e-value = 0.00) in 5 mammalian species (tenrec, little brown bat, mouse, rat, opossum) and in 2 tetrapods (green anole lizard and African clawed frog). In addition, one lower score, but potentially still significant hit (score = 330, e-value = 8e-27) was returned in guinea pig that had 84% sequence identity over 325 bp of the query. No other significant hits (scores ≤ 102, e-values ≥ 8e-19) were found in the other 27 vertebrate genomes or in any other species represented in the GenBank databases.

In tenrec and bat, *SPIN* was present in multiple full-length copies, which allowed the reconstruction of species-specific consensus sequences of 2,871 and 2,867 bp in length, respectively, both encoding a 601-aa hAT transposase. In the other species (opossum, mouse, rat, frog, and lizard), we could detect no more than two full-length, and often only partial copies,

precluding the reconstruction of reliable consensus sequences. Nonetheless, the longest *SPIN* copy from each of these species was used to construct a multiple alignment together with the consensus *SPIN* from bushbaby, tenrec and bat. The alignment revealed a strikingly high level of sequence identity over the entire length of *SPIN* elements (~2.9 kb), with an average of 96% pairwise nucleotide identity between any two species and 98% among the consensus sequences. The level of sequence identity was similarly elevated across both coding and non-coding regions (Figure 2.1a and Appendix B). Furthermore, the 16-bp terminal inverted repeats (TIR) of *SPIN* elements were all characterized by the same mismatch at position 4 (Figure 2.1b), suggesting that these elements all descend from the same active ancestral transposon carrying imperfect TIRs.



(a)

(b)

Figure 2.1 (A) Interspecific sequence identity across the entire length of full-length *SPIN* elements. The plot reflects the average sequence identity in non-overlapping bins of 50 nucleotides across a multiple alignment of the full-length *SPIN* consensus sequences from bat, tenrec and bushbaby, along with single copy sequences from frog, lizard and opossum. An average of 98% pairwise nucleotide identity is observed between the bat, tenrec and bushbaby consensus sequences and of 96% between any 2 sequences (range=84%-99%). The transposase ORF is depicted as a black rectangle and the terminal inverted repeats by arrowheads. (B) Multiple alignment of the 5' and 3' ends of full-length and MITE *SPIN* elements. The 16-bp terminal inverted repeats (TIRs) are boxed. All *SPIN* elements share the same imperfection at position 4 in their consensus (black arrowheads). The only exception is *SPIN_Xt*, the full-length *SPIN* element in frog for which we were only able to locate only a single partial copy. However, the same TIR imperfection is found in the consensus sequence of *SPIN_NA_5_Xt*, a MITE subfamily from frog. (Rodent = mouse/rat, Og = *Otolemur garnettii* (bushbaby), Et = *Echinops telfairi* (tenrec), Ml = *Myotis lucifugus* (bat), Xt = *Xenopus tropicalis* (frog), Ac = *Anolis carolinensis* (lizard), Md = *Monodelphis domestica* (opossum)).

2.2.2 Lineage-specific amplification of non-autonomous SPIN elements

A hallmark of DNA transposon evolution in eukaryotes is the proliferation of non-autonomous copies, also known as miniature inverted-repeat transposable elements (MITEs), which may greatly outnumber their full-length, autonomous partners (Feschotte and Pritham

2007; Feschotte et al. 2002). Typically, MITEs form homogeneous subfamilies that derive from the recurring trans-mobilization of a particular deletion derivative of a full-length copy (Feschotte and Pritham 2007; Feschotte et al. 2002). We found that each of the 8 species containing full-length *SPIN* copies also harbor distinct populations of related MITEs in their genome. A total of 12 major MITE subfamilies were detected (Table 2.1) and each subfamily was found to derive from a different deletion derivative of a full-length *SPIN* element (Appendix B & Figure 2.2). The presence of *SPIN* MITEs identified computationally was validated experimentally by PCR using oligonucleotide primer pairs internal to the TIRs of the *SPIN* elements (Figure 2.3 and see Materials and Methods). PCR products of the expected size were obtained from all species that were predicted to contain *SPIN*, but not from two mammalian species where *SPIN* was undetectable by Blast searches (human and Jamaican fruit bat). DNA sequencing of PCR products confirmed that they were identical or nearly identical to *SPIN* transposons found in the corresponding whole genome assemblies (GenBank accession numbers EU867495-EU867500 and FJ154100).

Table 2.1 For each species in which *SPIN* elements were discovered, the name of the element along with the length, copy number (of both full-length and fragmented elements) and average percent sequence divergence from the consensus sequence are shown. "NA" in the length column indicates that a consensus sequence could not be determined because too few full-length copies were identified. Likewise, average percent divergence could not be determined for full-length *SPIN* elements in rodents, opossum, frog and lizard due to low copy number and a high level of fragmentation. A total of 12 distinct MITE families, along with consensus sequences, could be identified in the 8 species. MITE families are italicized and denoted with the suffix NA for nonautonomous. A complete consensus sequence for *SPIN_NA_11_Ac*, a lizard-specific MITE family, could not be confidently reconstructed due to uncertainty in its central region. The copy number for this family was estimated based on counts of the 5' and 3' terminal regions, for which a reliable consensus could be reconstructed.

Species	TE Name	Length (bp)	Copy Number	Avg % Divergence
Mouse	**SPIN_Rodent**	**NA**	**<5**	**NA**
	SPIN_NA_10_Rodent	**225**	**34041**	**16.1**
Rat	**SPIN_Rodent**	**NA**	**<5**	**NA**
	SPIN_NA_10_Rodent	**225**	**32567**	**16.5**
Bushbaby	**SPIN_Og**	**2836**	**7145**	**7.2**
	SPIN_NA_1_Og	**225**	**8480**	**8.6**

Table 2.1 - *continued*

	SPIN_NA_2_Og	**80**	**17498**	**11.9**
Tenrec	**SPIN_Et**	**2871**	**13963**	**9.4**
	SPIN_NA_1_Et	**224**	**52551**	**10.4**
	SPIN_NA_6_Et	**487**	**32824**	**9.4**
Bat	**SPIN_Ml**	**2867**	**2806**	**3.9**
	SPIN_NA_7_Ml	**212**	**21198**	**3.2**
	SPIN_NA_8_Ml	**192**	**3693**	**5.5**
	SPIN_NA_9_Ml	**311**	**10638**	**3.1**
	SPIN_NA_10_Ml	**223**	**11984**	**3.3**
Opossum	**SPIN_Md**	**NA**	**<5**	**NA**
	SPIN_NA_3_Md	**192**	**3671**	**3.9**
	SPIN_NA_4_Md	**718**	**1136**	**5.7**
Frog	**SPIN_Xt**	**NA**	**<5**	**NA**
	SPIN_NA_5_Xt	**186**	**3992**	**5.2**
Lizard	**SPIN_Ac**	**NA**	**<5**	**NA**
	SPIN_NA_11_Ac	**273**	**12138**	**9.1**

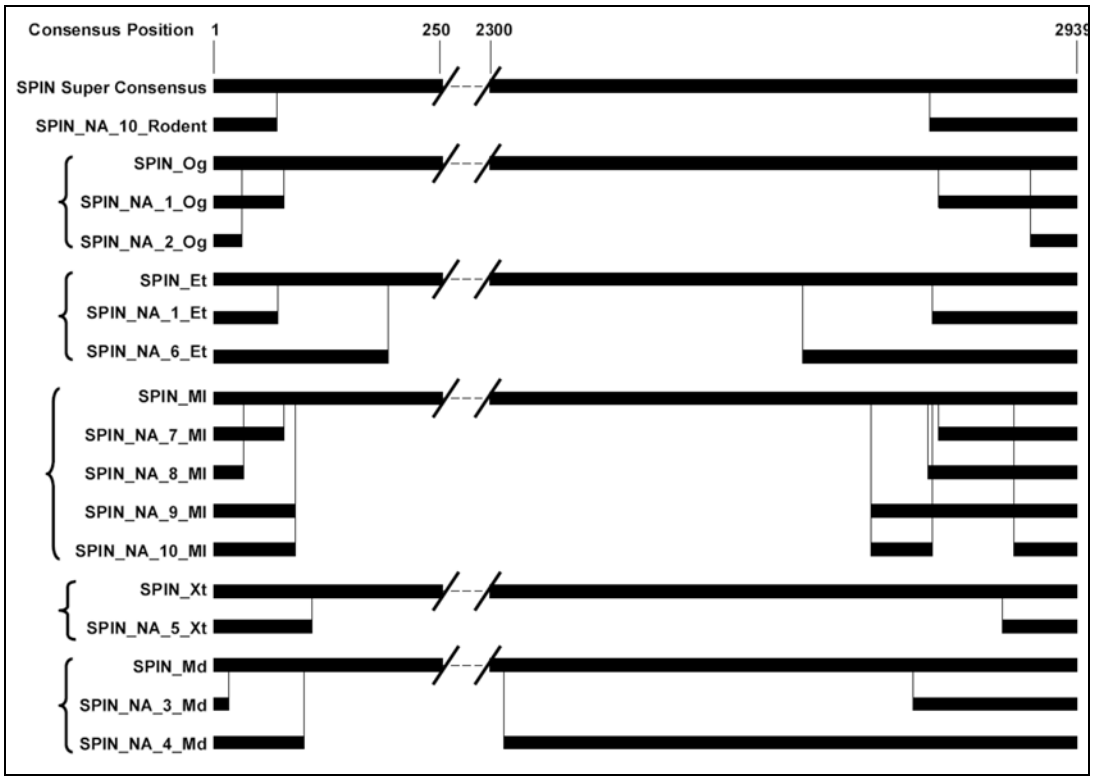


Figure 2.2 Schematic of regions of shared identity between consensus sequences of full-length *SPIN* transposons and their derived MITEs in each species. MITEs are denoted by the suffix 'NA' (for nonautonomous). The figure shows that the MITE families from different species have distinct deletion breakpoints with their cognate full-length *SPIN* transposon, indicating that each originated independently from a distinct deletion derivative. Species abbreviations: Rodent = mouse/rat, Og = *Otolemur garnettii* (bushbaby), Et = *Echinops telfairi* (tenrec), Ml = *Myotis lucifugus* (bat), Xt = *Xenopus tropicalis* (frog), Ac = *Anolis carolinensis* (lizard), Md = *Monodelphis domestica* (opossum).

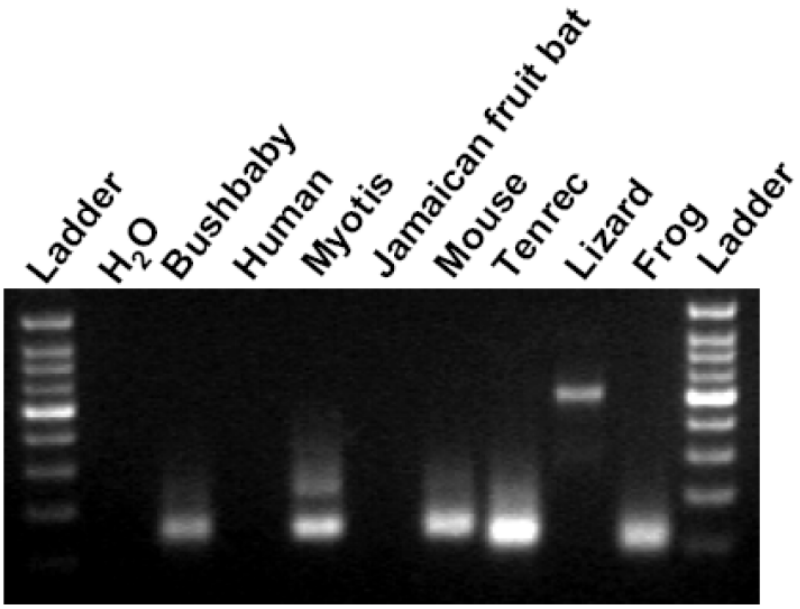


Figure 2.3 Experimental verification of the presence or absence of *SPIN* transposons. PCR fragments of expected sizes were obtained in all species (or a close relative) where *SPIN* elements were identified computationally. A PCR product of the expected size (not shown here) was also obtained with DNA from the opossum, *Monodelphis domestica*. DNA sequencing of cloned PCR products for each species confirmed that they represent distinct *SPIN* family members (see Materials and Methods).

While the consensus sequences for each MITE subfamily differ in size, their shared regions display a level of interspecific pairwise sequence identity (>91%) comparable to full-length *SPIN* elements (Table 2.1). They also display the same imperfect TIRs (Figure 2.1b). Within each species the MITE subfamilies achieved very high copy numbers, ranging from ~4,000 in frog to ~99,000 copies in tenrec (Table 2.1). Thus, while full-length *SPIN* elements from different species are indistinguishable in terms of their sequence and structure, they have colonized the 7 species lineages with differential success and given rise to structurally distinct MITE subfamilies in each of these lineages.

2.2.3 Evidence for the horizontal introduction of SPIN elements

The level of sequence identity of *SPIN* transposons among such widely diverged tetrapods is exceptional, being greater than some of the most conserved protein-coding genes in vertebrates (e.g. RAG-1) (Hugall et al. 2007) and comparable to the so-called ultraconserved elements (Bejerano et al. 2004). Such levels of sequence identity can be explained by one of

two alternatives: either *SPIN* elements have been vertically inherited from the last common ancestor of these species (~360 Myr ago (Hedges and Kumar 2004)) and preserved by intense purifying selection in these lineages; or full-length *SPIN* progenitors were introduced horizontally in these lineages and subsequently spread within each genome.

Several observations indicate that a scenario of vertical acquisition of *SPIN* elements is untenable. Under this hypothesis, the patchy taxonomic distribution of *SPIN* would require that these elements were lost repeatedly during tetrapod evolution (at least 14 times independently (Figure 2.4)), while being maintained active in a subset of lineages. This scenario would also imply that some ancestral *SPIN* copies have been retained at orthologous genomic positions in some of these species or in the sister taxa represented in the databases. Such ancient transposon fossils are generally well preserved in the genomes of eutherians owing to their relatively slow substitution rates (Waterston et al. 2002; Yang et al. 2004; Pace and Feschotte 2007). However, with the exception of mouse and rat, where nearly all *SPIN* elements were found at orthologous genomic positions and therefore must have inserted prior to the divergence of these rodents (Figure 2.4), none of the *SPIN* elements from one species were present at orthologous positions in any of the other species for which complete or nearly complete genome sequences are available. Moreover, we could readily identify *SPIN* copies present only in one species (bushbaby, bat or tenrec), but precisely absent at orthologous position in species representing other mammalian orders (Figure 2.5). These results strongly indicate that *SPIN* amplification occurred independently in each of these lineages and that it postdates the radiation of these mammalian orders. Together, these data are inconsistent with a scenario of ancient origin followed by vertical persistence of *SPIN* activity throughout tetrapod evolution.

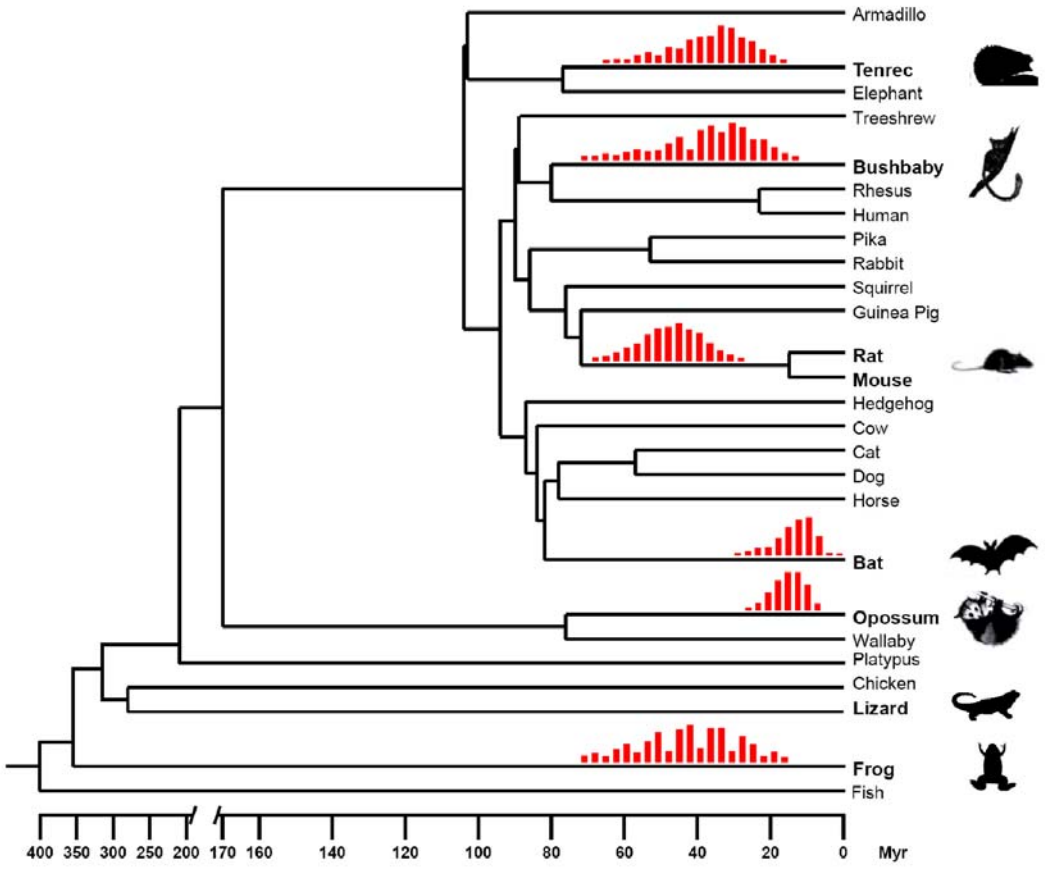


Figure 2.4 Species distribution and timing of amplification of *SPIN* transposons. The tree depicts the phylogenetic relationship and divergence times of the vertebrate species with complete or nearly complete genome sequences currently available (Hugall et al. 2007; Murphy et al. 2007). The species harboring *SPIN* transposons are in bold. The timing of *SPIN* amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all *SPIN* MITE subfamilies found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin (see Materials and Methods in supporting information). The age span is not shown for lizard because the neutral substitution rate is undetermined for this species. However, we note that the level of sequence divergence of *SPIN* in lizard is similar to those observed in bushbaby and tenrec (see Table 2.1).

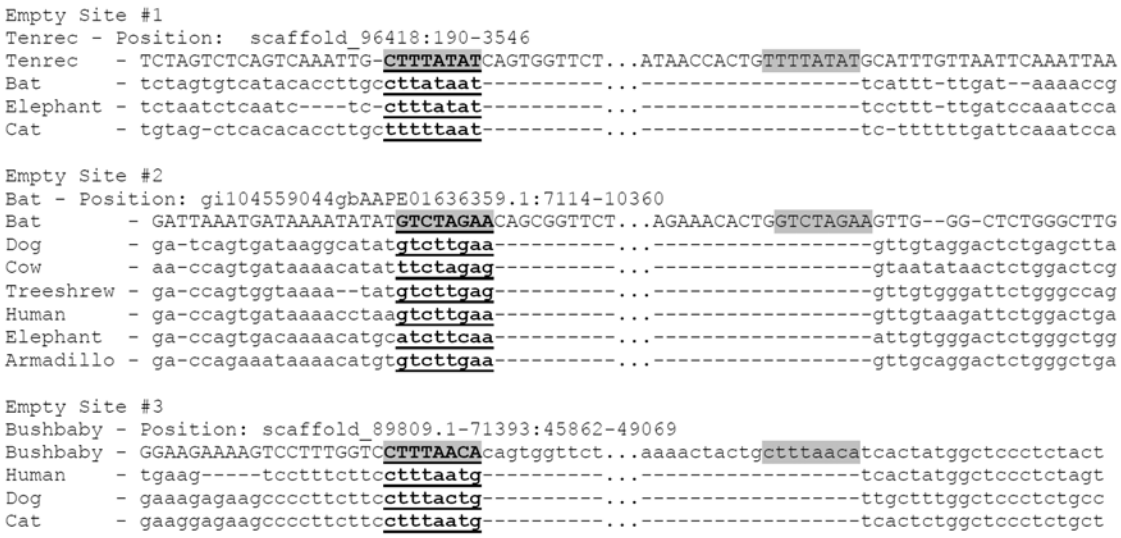
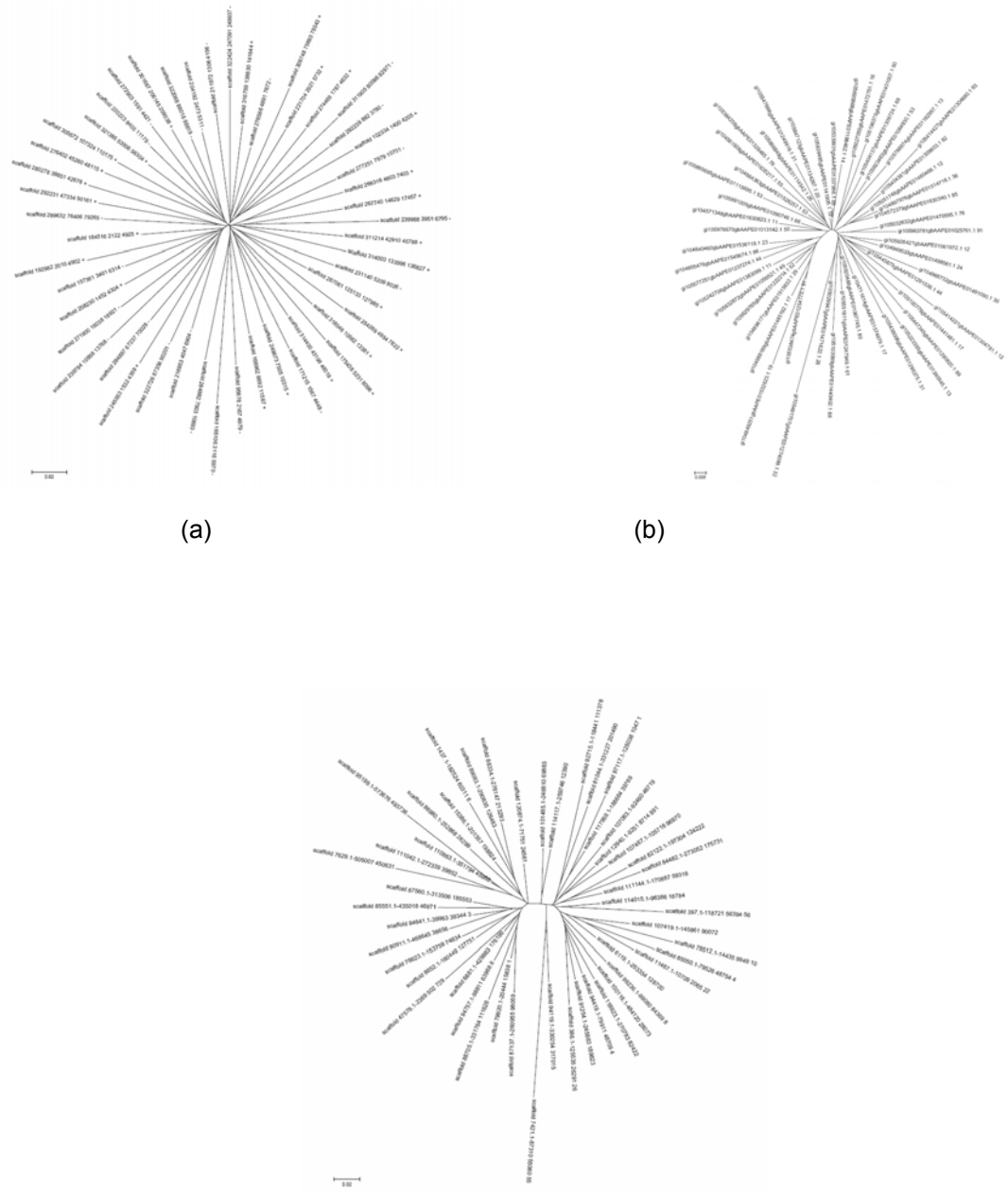


Figure 2.5 *SPIN* orthologous "empty" sites. Multiple alignments showing the presence (top line, reference species) or absence of *SPIN* elements at orthologous loci in diverse mammalian species. In all cases, we found both TSD copies (underlined) and the individual *SPIN* insertion in the reference species but only one copy of the TSD and no additional sequences in the comparison species.

In addition, several lines of evidence allowed us to rule out that *SPIN* elements, as a whole, have evolved under purifying selection in the lineages examined. First, neighbor-joining phylogenies of *SPIN* elements mined from each species revealed a star-like topology indicative of a single burst of transposition followed by accumulation of discrete mutations along each branch (Figure 2.6). This idiosyncratic pattern of evolution is consistent with neutral evolution and is typical of DNA transposons (Hartl et al. 1997; Feschotte and Pritham 2007). Furthermore, we found no significant difference in the level of sequence identity among full-length *SPIN* copies at the first, second or third codon positions of the transposase ORF ($p > 0.05$, $\chi 2$ test) and no signature of purifying selection acting on *SPIN* transposase sequences since their divergence from their ancestral consensus sequence (i.e. Ka/Ks not significantly smaller than 1, see Materials and Methods). The single exception to this pattern is a peculiar *SPIN* transposase ORF found at orthologous position in mouse and rat (chr5:130,377,377-130,379,669 in the mm8 mouse genome assembly, see Figure 2.7) that has apparently evolved under strong evolutionary constraint since the divergence of the two murine species (Ka/Ks = 0.10; significantly less than 1, $p < 0.0001$). The non-coding regions of the corresponding *SPIN*

element are still recognizable, but they have suffered secondary transposable element insertions and extensive mutational decay (Figure 2.7). Based on mouse transcriptome data (e.g. full-length cDNA clones BC137610 and AK010551, see Figure 2.7), this *SPIN* transposase appears to be expressed from a far-upstream promoter in fusion with flanking exons encoding a CHCH domain (PFAM06747). The resulting putative chimeric protein (756 aa) might have been 'exapted' for a cellular function in the murine lineage, akin to the SETMAR protein of primates, which results from the fusion of a *mariner* transposase with flanking exons of a SET domain gene (Cordaux et al. 2006b). Notwithstanding this possible case of transposase exaptation, we can reject the hypothesis that the extreme level of *SPIN* sequence identity among widely diverged tetrapods reflects the systematic action of purifying selection. Thus, the only plausible scenario is that active and nearly identical *SPIN* elements were introduced horizontally, and relatively recently, into several tetrapod species and subsequently spawned different waves of *SPIN* amplification along these species lineages.



(a) (b)

(c)

Figure 2.6 Phylogenetic relationship among members of the same *SPIN* family. Neighbor-joining trees of (A) a family of full-length elements from tenrec (*SPIN_Et*), (B) a MITE family from bat (*SPIN_NA_9_Ml*) and (C) a MITE family from bushbaby (*SPIN_NA_1_Og*). For (A), all 49 full-length copies of *SPIN_Et* were included, while for both (B) and (C), 50 copies of each family were randomly extracted from each genome and a multiple alignment was created using ClustalX (v. 1.83) . The neighbor-joining tree was constructed using MEGA 3.1 (1,000 bootstrap replications). The star-like topology of each tree and the lack of subfamily structure are indicative of a single burst of transposition followed by a pattern of idiosyncratic evolution along each branch.

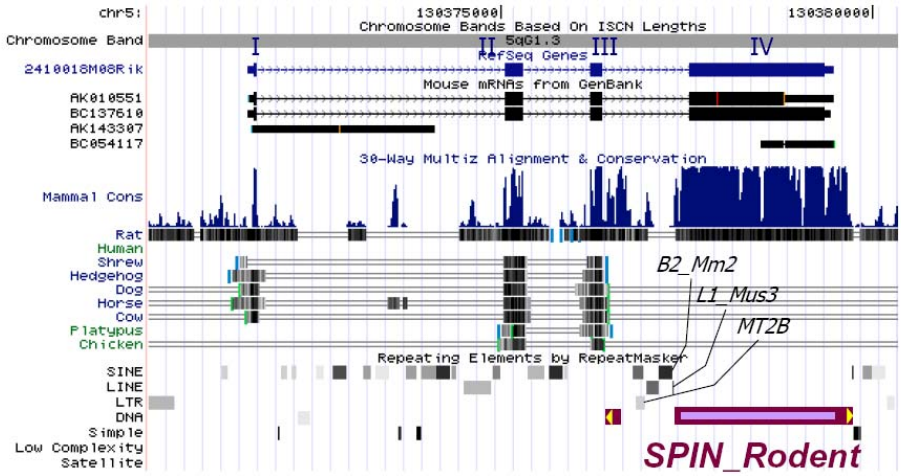


Figure 2.7 Murine-specific chimeric gene derived from *SPIN* transposase. This snapshot was extracted from the annotation of the mouse July 2007 (mm8) genome assembly implemented within the UCSC Genome Browser. The chromosomal region displayed is chr5:130,370,270-130,380,588. The dark blue rectangles at the top represent the four exons of RefSeq gene 2410018M08Rik (also known as NM_183088.2). Exons I-III encode a CHCH domain (pfam06747), while exon IV is entirely derived from a *SPIN* transposase and, accordingly, is predicted to encode a hATc domain (pfam05699). The black rectangles depict corresponding exons identified in mouse cloned cDNAs deposited in Genbank. Below is the conservation track, which reflects the alignment of this mouse genomic region with other vertebrate genomes (species listed on the left). Exons I-III are present and conserved in most other mammals (but apparently not in human) and in chicken, while exon IV is absent from these species, except rat where all four exons are present and highly conserved. The last set of annotations at the bottom depicts the position of repeats given by Repeatmasker. The relative position of *SPIN_Rodent* is shown by the purple bar, with the TIRs marking the boundaries of the element shown as yellow arrowheads and the transposase ORF as a violet horizontal line. The Repeatmasker annotation shows that the 5' non-coding region of *SPIN_Rodent* has suffered at least consecutive insertions of three other mobile elements (MT2B ERVL, L1_Mus3 LINE, B2_Mm2 SINE). These insertions did not disrupt the ORF, but likely caused the immobilization of the *SPIN_Rodent* element.

2.2.4 Timing of SPIN colonizations

To gain further insight into the timing of the transfers, we derived an estimate of the time elapsed since the amplification of *SPIN* families in each of the host species lineages. Since the elements have evolved neutrally since their insertion, the age of individual insertions can be approximated by measuring the sequence divergence from the ancestral consensus sequence and by applying a neutral substitution rate characteristic of the species lineage (Waterston et al. 2002; Yang et al. 2004; Pace and Feschotte 2007). To estimate the neutral substitution rate in the lineages of bushbaby, murine rodents, bat and tenrec, we retrieved, aligned and compared

a large and diverse set of ancient transposable elements present at orthologous genomic positions in human, which is known to have diverged from each of these species approximately 80, 89, 94 and 104 mya, respectively [see Materials and Methods]. For the opossum lineage, we used ancient repeats found at orthologous positions in opossum and wallaby, which diverged ~76 mya. Based on the number of substitutions per million years, we infer that both full-length and MITE *SPIN* families amplified within a fairly narrow evolutionary window (~31-46 mya) in the lineages of rodents, bushbaby, tenrec and frog, consistent with a single wave of germline infection of diverse tetrapods (Figure 2.4). The burst of *SPIN* amplification seems more recent in the bat and opossum lineages, ~15 mya (Figure 2.4), which may indicate more recent *SPIN* transfers in these lineages or a delay in their amplification following the initial horizontal introduction.

2.3 Discussion

Our analysis provides the first unequivocal evidence for the HT of a DNA transposon into distant tetrapods, including species from five distinct mammalian orders. These findings are startling given the barriers apparently opposing the entry of exogenous DNA into the sequestered germline of animals (Silva et al. 2004). The few convincing cases of HT that have been documented generally involve transfers between fairly closely related taxa, such as P and copia elements among drosophilids (Silva et al. 2004; Jordan et al. 1999), MULE transposons between grasses (Diao et al. 2006) or the concurrent germline infiltrations of PTERV1 retroviruses in chimpanzee and gorilla (Yohn et al. 2005). Some DNA transposons, such as *mariner* elements, are apparently able to cross wider evolutionary distances, in part because their transposition does not seem to require specific host factors (Silva et al. 2004; Robertson 2002; Hartl et al. 1997; Yoshiyama et al. 2001). *SPIN* is remarkable in that it entered the germline and reached high copy numbers in placental, marsupial, reptilian and amphibian species. This is consistent with the ability of hAT superfamily transposons to jump efficiently in a wide range of heterologous species and conditions (Weil and Kunze 2000; Evertts et al. 2007; Kawakami 2007).

The apparent overlap in the timing of *SPIN* amplification in the different lineages (Figure 2.4) and the fact that the *SPIN* consensus sequences are phylogenetically equidistant to each other (Figure 2.1a, Appendix B and Figure 2.8) suggest that different tetrapod species were infected at close evolutionary time points by essentially the same active element. Presumably, this element originated from the same ancestral source species. At the moment, the taxonomic identity of the source species cannot be identified because *SPIN*-encoded transposases revealed no more than ~45% amino acid similarity with other eukaryotic hAT transposases found in the databases (Figure 2.8). It is also not possible to distinguish whether each tetrapod species acquired *SPIN* independently from the same exogenous source (e.g. a common prey or parasite) or if *SPIN* was acquired once by a tetrapod species from an exogenous source and then spread by HT between tetrapod species. The mechanisms underlying such recurrent HTs remain to be clarified, but we note that DNA viruses are increasingly recognized as potential intermediates for HT of mobile elements between widely diverged animals (Piskurek and Okada 2007; Fraser et al. 1983; Friesen and Nissen 1990; Jehle et al. 1998). Recently Piskurek and Okada (Piskurek and Okada 2007) reported on the transfer of a snake retroposon into the genome of taterapox virus (TATV), a poxvirus that infects West African rodents. Vertebrate poxviruses are good candidates as *SPIN* vectors because many have the ability to infect a broad range of species and cell types and some, such as smallpox virus, have been implicated in zoonosis (McFadden 2005). Interestingly, several of the mammals harboring *SPIN* (i.e. bats, opossum, murine rodents) are notorious reservoir species for diverse viruses, including poxviruses. However our current view of the taxonomic distribution of *SPIN* is heavily biased by the sample of species available in the databases. To gain insights into the origin and evolutionary history of *SPIN*, future experiments shall focus on a systematic exploration of the taxonomic distribution of these elements.

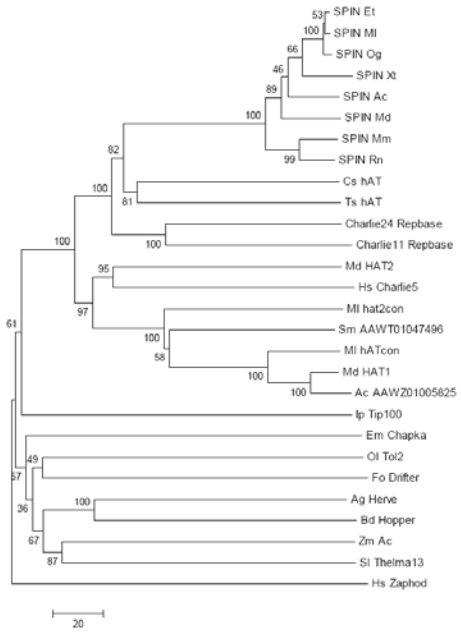


Figure 2.8 Phylogenetic relationship of *SPIN* transposase within the eukaryotic hAT superfamily. The neighbor-joining tree was constructed using MEGA 3.1 (1,000 bootstrap replications) (2) based on a multiple alignment of the amino acid sequence of the *SPIN* transposases and various representatives of the hAT superfamily of transposases. Each sequence is identified by the species initials (abbreviations shown at the bottom of the tree), followed by the name of the corresponding transposon, if previously described. Other hAT elements newly identified in this study are designated by the suffix 'hAT'. The hAT transposases most closely related to *SPIN* are found in the chordate Ciona savignyi (hAT_Cs) and the roundworm Trichinella spiralis (hAT_Ts). However, each of these transposases has less than 45% amino acid similarity to *SPIN*.

Despite these outstanding questions, our data provide evidence that HT has contributed significantly to diversifying and shaping the genomes of mammals and other tetrapods. With copy numbers per haploid genome ranging from 4,000 to nearly 100,000, *SPIN* ranks, to the best of our knowledge, as the most successful DNA transposon family ever reported in any species. DNA transposons have been shown to cause both large- and small-scale genomic rearrangements due to ectopic recombination or aberrant transposition events and they have also contributed to the creation of new genes (Feschotte and Pritham 2007). Here we discovered an apparently functional, murine-specific chimeric gene derived from a *SPIN*

transposase (Figure 2.7). Thus, there is little doubt that *SPIN* amplification not only added megabases of DNA to the genomes, but it also promoted lineage-specific changes in chromosomal architecture and fueled evolutionary innovation.

2.4 Materials and Methods

2.4.1 Identification of SPIN elements

The sequence of *SPIN_NA_2_Og*, also known as PMER1 in Repbase (Jurka et al. 2005), was used as a query in local Blastn searches (v. 2.2.14) (Altschul et al. 1997) against the bushbaby whole genome shotgun (WGS) assembly (otoGar1) downloaded from the UCSC Genome Browser (http://genome.ucsc.edu) (Kent et al. 2002). Sequences matching the 5' and 3' terminal regions of *SPIN_NA_2_Og* separated by more than 2,000 bp were extracted and used to reconstruct the consensus sequence for the full-length *SPIN_Og* element. The *SPIN_Og* consensus sequence was used as a query in Blastn searches against all NCBI Blast databases (including trace archives) to determine the presence or absence of *SPIN* elements in other species. The following Blastn parameters were used: gap existence penalty: 6, gap extension penalty: 5, penalty for nucleotide mismatch: -5, reward for nucleotide match: 4. *SPIN* elements were considered present within a species if the Blastn score (bits) for a high scoring pair (HSP) was greater than or equal to 1718, a value corresponding to 80% of the query sequence matching a sequence within the species database with 90% identity. We chose to use scores rather than e-values since the later are dependent upon the size of the queried database, which varies between species. However, we note that in all species where an HSP with a score greater than 1718 was obtained, at least one such HSP had an e-value of 0.00. *SPIN* elements were considered to be absent from a species if the best HSP had a score less than or equal to 158, corresponding to 25% of the query sequence matching a sequence within the species database with 90% identity. When *SPIN* elements were found within a genome, consensus sequences were constructed for each family using a simple majority rule based upon a multiple alignment of at least 20 copies. *SPIN* consensus sequences have been deposited in Repbase Update (Jurka et al. 2005). We also derived a *SPIN* "super-consensus" by creating a

multiple alignment of full-length *SPIN_Ml*, *SPIN_Et* and *SPIN_Og* consensus sequences. To determine copy numbers and average percent divergence of individual *SPIN* families, we used the respective consensus sequences to mask the corresponding genome sequence using RepeatMasker v. 3.1.5 (Smit et al. 1996-2004). We note that some of the MITE consensus sequences used have been published previously in Repbase. These include: *SPIN_NA_10_Rodent* (URR1A), *SPIN_NA_2_Og* (PMER1), *SPIN_NA_5_Xt* (URR1_Xt), *SPIN_NA_3_Md* (URR1A_Mdo), *SPIN_NA_4_Md* (URR1B_Mdo), *SPIN_NA_7_Ml* (nhAT4a_ML), *SPIN_NA_8_Ml* (nhAT4b_ML), *SPIN_NA_9_Ml* (nhAT5a_ML) and *SPIN_NA_10_Ml* (nhAT5b_ML). All other consensus sequences were derived during this work.

2.4.2 Tests for purifying selection

To examine the pattern of evolution of *SPIN* coding sequences, we retrieved the least degraded transposase sequences from bat, tenrec and bushbaby, aligned them to their respective consensus sequences (which offers a theoretical approximation of their ancestral sequence), removed codons containing obvious nonsense and frameshift mutations, and computed pairwise Ka/Ks ratios with Mega 3.1 (Kumar et al. 2004) using the model "Codon: Nei-Gojobori method with the Jukes-Cantor correction". P-values were calculated using the codon-based Z-test with the "Codon: Modified Nei-Gojobori method with the Jukes-Cantor correction" model and 500 bootstrap replications.

2.4.3 Calculation of neutral substitution rates in mammalian lineages colonized by SPIN

Neutral substitution rates have not been published for each of the species where *SPIN* elements are found. Therefore, we devised a method that uses alignments of ancestral DNA transposons, which are assumed to have evolved neutrally since their insertion (Waterston et al. 2002; Yang et al. 2004) as a proxy to derive the neutral substitution rate in the species lineages invaded by *SPIN*. For each target species, we identified over 400 individual insertions (range = 404-1,469) from 12 different DNA transposon families present at orthologous loci in human using a program called OrthoBlast (Pace & Feschotte, unpublished). Orthologous TE copies were aligned pairwise and to their ancestral consensus sequence (published in Repbase (Jurka

et al. 2005)) and gaps were removed to produce a 3-way alignment (target species/human/consensus). All alignments were concatenated to derive an overall substitution rate. For the opossum, we used 846 DNA transposons from 6 different families present at orthologous loci in opossum and wallaby. The overall percent divergence of all of the orthologous TE copies was calculated with the Kimura 2-parameter correction (Kimura 1980). A neutral substitution rate was then inferred by dividing the percent divergence by the estimated time elapsed since the split of the two compared species. The divergence times that were used to infer neutral substitution rates were derived from a survey of the most current literature (Janecka et al. 2007; Murphy et al. 2007; Poux et al. 2006; Poux et al. 2005; Delsuc et al. 2004; Steiper and Young 2006; Bininda-Emonds et al. 2007; Springer et al. 2003; Yoder et al. 2003) (5-13). These divergence times, along with the calculated rates in substitutions/myr, are: human-tenrec: 104 myr (2.9173 x 10-9), human-bat: 94 myr (2.6920 x 10-9), human-rodents: 89 myr (3.5411 x 10-9), human-bushbaby: 80 myr (2.9590 x 10-9), opossum-wallaby: 76 myr (3.2113 x 10-9).

We note that our estimate of the neutral substitution rate for the mouse lineage is significantly lower than the one published previously (4.5 x 10-9 substitutions/myr) (Waterston et al. 2002). However, this discrepancy can be largely explained by the fact that the previous study used a divergence time between rodents and humans of 75 myr, while the most recent literature (Janecka et al. 2007; Murphy et al. 2007; Poux et al. 2006; Poux et al. 2005; Delsuc et al. 2004; Steiper and Young 2006; Bininda-Emonds et al. 2007; Springer et al. 2003; Yoder et al. 2003) points to a deeper divergence date (we used 89 mya based on a consensus of the literature). Because it was not possible to find a large number of orthologous repeats present either in frog or lizard and another species with known divergence time, we used a neutral substitution rate previously published in frogs (1.2270 x 10-9) (Crawford 2003) to infer the age of *SPIN* elements found in Xenopus, but were unable to infer the age of *SPIN* in lizard.

2.4.4 PCR and sequencing of non-autonomous SPIN elements

The presence of non-autonomous *SPIN* elements was verified by PCR/sequencing in all the species where they were identified computationally (or in a close relative): bushbaby (*Otolemur garnettii*), Ridley's bat (*Myotis ridleyi*), common tenrec (*Tenrec ecaudatus* – SMG-15088; S.M. Goodman's collection), mouse (*Mus musculus*), Anole lizard (*Anolis carolinensis*) and Western clawed frog (*Xenopus tropicalis*), opossum (*Monodelphis domestica*). Two species where *SPIN* elements were not found computationally were added as negative controls: human (*Homo sapiens*) and the Jamaican fruit bat (*Artibeus jamaicensis*).

Non-autonomous *SPIN* elements were amplified in all species but opossum using the following PCR primers: NA-F 5'CGA ACG ACC CTT TCA CAG G (position 41-59 of the super consensus, (Appendix B) and NA-R 5'CAG TTC CTC ATG TTG TGG TGA C (position 2878-2899). The primers used for the opossum are: NA-Fmdo 5'GGT CGC CTA AAG CCA TCG (position 60-77) and NA-Rmdo 5'GGT CGC CTA AAG CCA TCG. PCR was carried out with an initial denaturation step of 2 min at 94 °C followed by 30 cycles of 1 min at 94 °C, 30 s at 53 °C, and 30 s at 72 °C. PCR mix was: Buffer (5x) 5 ul, MgCl 2 (25 mM) 2 ul, dNTP (10 mM) 0.5 ul, NA-F (10uM) 1ul, NA-R (10 uM) 1ul, Taq (GoTaq, Promega) 1.25 U, DNA 30ng, H2O up to 25 ul. PCR products were cloned into the pCR2.1-TOPO cloning vector (Invitrogen) and one randomly selected clone of each species was sequenced.

CHAPTER 3

CAPTURE OF DNA VIA DOUBLE-STRAND BREAK REPAIR IS A COMMON MECHANISM OF GENOMIC DUPLICATION IN VERTEBRATES

3.1 Introduction

Environmental agents and normal cellular metabolic processes produce DNA double-strand breaks (DSBs) that can lead to cell death if not repaired (Haber 2000). Eukaryotic cells have evolved DSB repair mechanisms that can be classified into two broad categories. The first is the canonical homologous recombination (HR) pathway. HR uses long stretches of homology between the flanking sequences at the site of breakage and the homologous chromosome or sister chromatid to repair DSBs perfectly, leaving no evidence that a break ever occurred. The second category includes three separate mechanisms, each of which repairs the break imperfectly. The first two mechanisms, non-homologous end joining (NHEJ) and single strand annealing (SSA), repair the DSB without the use of a repair template and typically create deletions at the site of the lesion. The third mechanism, synthesis-dependent strand annealing (SDSA), is a form of homologous recombination that uses an ectopic template sequence to repair the DSB (Figure 3.1) (Haber 2000).

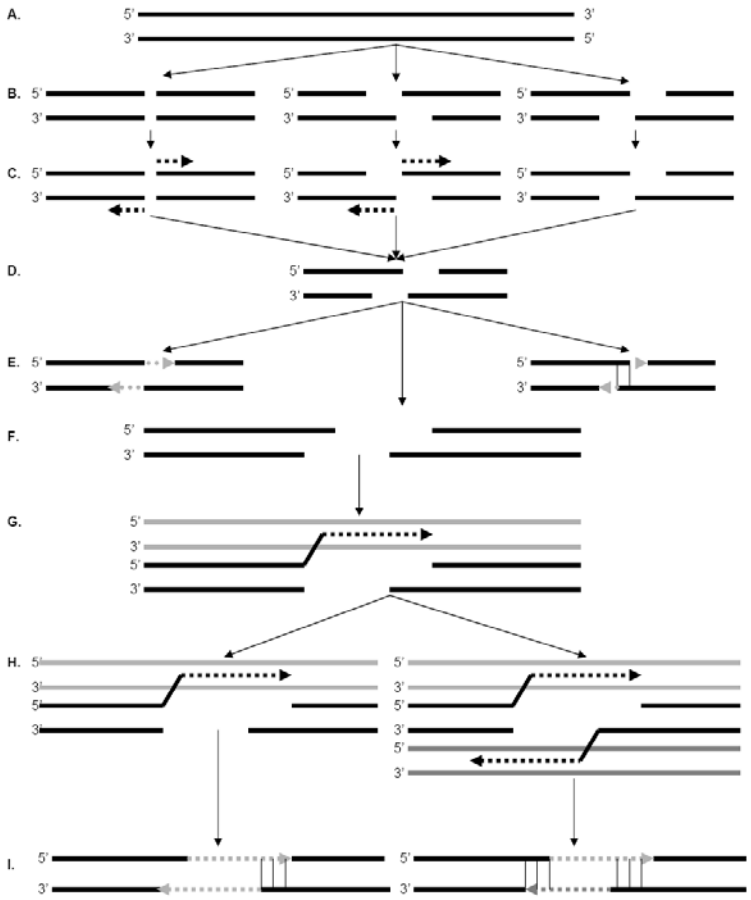


Figure 3.1 DSB repair pathways. (a) A DNA molecule suffers a double-strand break. (b) The DSB can result in either blunt ends (left), 5' overhanging ends (center) or 3' overhanging ends (left). (c) After the break occurs, exonuclease activity (dotted arrow) creates 3' overhanging ends at the site of the lesion by removing nucleotides from blunt breaks (left) or breaks with 5' overhanging ends (center). (d) After the exonuclease activity ceases, the resulting breakpoint have 3' overhangs at both ends of the lesion. (e) Three major pathways are available to repair the lesion. First, the break can be repaired by homologous recombination using a homologous chromosome or sister chromatid, resulting in a perfect repair (left). Second, if a region of microhomology exists between the two 3' overhanging ends, the NHEJ pathway can repair the break, typically causing a small deletion (right). (f) The third pathway to repair the lesion is SDSA. (g) One 3' overhanging end invades another DNA molecule, annealing to a sequence that is complementary, and repair synthesis begins. (h) The other 3' overhanging end may also invade a DNA molecule and begin repair synthesis. (i) After repair synthesis is completed and the new strand has dissociated from the template strand, it anneals with the other end of the initial lesion at a complementary region, creating the duplication of one template sequence (left). If both 3' ends of the lesion used different template strands for repair synthesis, both strands will anneal at a complementary region, resulting in the duplication of 2 different template sequences (right).

Previous empirical studies have provided a detailed characterization of the breakpoints produced by imperfect DSB repair mechanisms in eukaryotic cells. In vivo and ex vivo systems,

designed to track the fate of experimentally induced DSBs in yeast, fly, plant and mammalian cells, have shown that imperfect repair is often accompanied by the insertion of extra, 'captured' DNA at the breakpoint (Nassif et al. 1994; Moore and Haber 1996; Gorbunova and Levy 1997; Liang et al. 1998; Salomon and Puchta 1998; Lin and Waldman 2001; Ricchetti et al. 1999; Yu and Gabriel 1999; Richardson and Jasin 2000a). Several studies have found that this captured DNA is a duplication of sequences that have homology with the experimentally induced breakpoints such as (1) a different chromosome (Varga and Aplan 2005; D'Anjou et al. 2004), (2) extra chromosomal molecules such as plasmids or mitochondrial DNA (Yu and Gabriel 1999; Ricchetti et al. 1999; Liang et al. 1998; Decottignies 2005), (3) cDNA copies of retrotransposons (Teng et al. 1996; Yu and Gabriel 1999) or (4) a nearby sequence on the same chromosome (Phillips and Morgan 1994; Gorbunova and Levy 1997). One such mechanism for "capturing", and thereby duplicating a sequence during DSB repair, is the SDSA pathway.

Repair-mediated duplications (RD) may occur via SDSA when nucleotides on one of the overhanging 3' ends of the DSB anneal with a complementary sequence that serves as a template for synthesis. After synthesis is completed, the invading strand is displaced and anneals to the other resected 3' end of the DSB, resulting in conservative duplication of the template sequence. Additionally, both ends of the break are free to anneal with different templates, initiate repair synthesis, then re-anneal with each other. In this case, the duplication of two different templates occurs at the site of breakage (Figure 3.1) (Haber 2000).

In order to assess the impact of duplications resulting from imperfect DNA repair upon vertebrate genome evolution, we sought to computationally identify duplications whose breakpoints bear the previously characterized hallmarks of imperfect repair events in 10 vertebrate genomes. Our analysis includes primarily primates (human, chimp, orangutan, Rhesus macaque), but also chicken, zebrafish and other mammals (mouse, rat, dog, cow). We discovered 824 RDs in the human genome, 15 of which were found to be specific to the human

lineage, as they are absent from the chimp, orangutan, Rhesus macaque and marmoset. We confirmed experimentally that one of the human-specific RDs remains polymorphic in the general population. Lineage-specific RDs were found in all genomes for which a closely related ancestor was available for validation. Thus, RDs are a previously under appreciated force shaping eukaryotic genome architecture and contributing to structural genomic variation in humans and other vertebrates.

3.2 Results

3.2.1 Identification of Potential RDs

In order to find duplications whose breakpoints bear the previously characterized hallmarks of imperfect repair events, but not those of other known duplication mechanisms (such as retroposition), we developed a novel computational approach that capitalizes upon the repetitive nature of eukaryotic genomes. In humans and other primates, about 45% of the genome is composed of interspersed repeats that derive from the activity of a limited number of transposable element (TE) families (Lander et al. 2001). Since the derived ancestral consensus sequence for any family of TEs is known, the insertion of a sequence within a TE can be found by locating fragments of annotated TEs that are separated by an intervening sequence that is not alignable with the consensus sequence (Figure 3.2).

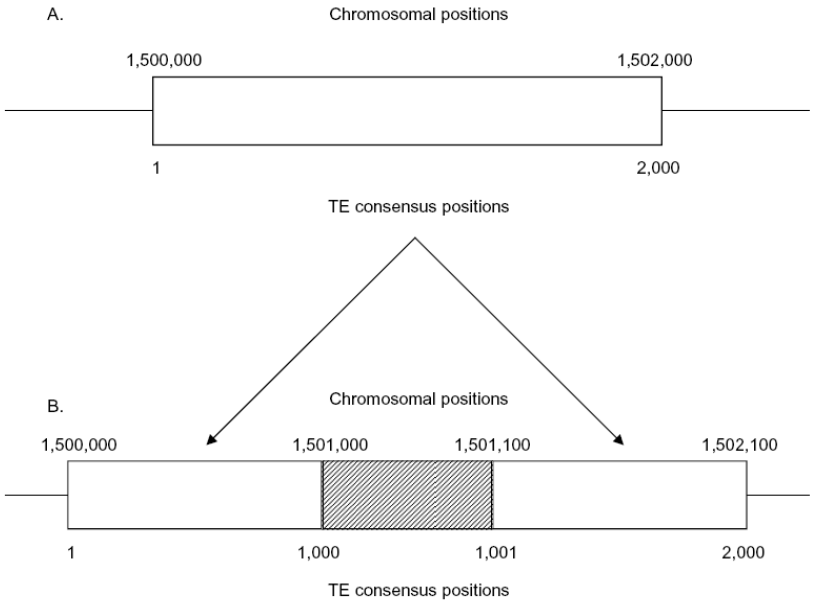


Figure 3.2 Identification of interrupted TEs. (a) Annotation of an uninterrupted TE (rectangle) along a chromosome. The chromosomal location is shown above the rectangle, while the positions of the TE consensus sequence are shown below. (b) Annotation an interrupted TE. After the insertion of a new sequence E, the 2 TE fragments are separated along the chromosome.

Drawing upon the wealth of TE annotation available, our computational pipeline began with a Perl script that parsed the human RepeatMasker (rmsk) annotation files (hg18 assembly, http://genome.ucsc.edu) to identify TEs that were interrupted by an insertion of at least 50 bp. The first pass returned over 492,000 such instances. The majority of these insertions were identified as segmental duplications, transposon insertions, processed retrogenes or LINE1 3' transduction events and were filtered from the dataset (see Methods). We next excluded inserted sequences not identified as any of the above types of duplications, but found at more than one other place in the genome or where a second, parental copy of the insertion could not be confidently identified. Finally, we removed any cases where the inserted sequence and putative template sequence were within 50 bp of each other to rule out the possibility of tandem duplications, which are typically formed by a mechanism other than DSB repair (Levinson and Gutman 1987). The remaining dataset of potential RDs included 1,136 interrupted TEs that had

suffered the insertion of a sequence found at only one other location in the genome, each of which may represent captured DNA segments duplicated via DSB repair.

Strikingly, of the 1,136 potential RDs in our dataset, 824 of the donor sequences were within 5 kb of the acceptor, with 753 (66%) separated by less than 3 kb. The remaining 312 duplicons were separated by more than 5 kb (n=58, max= 122 mb) or were located on different chromosomes (n=254). A graph of the distance between donors and acceptors on the same chromosome demonstrates two marked peaks in the distance separating the duplicons, with the first near 1,200 bp and the other at > 5,000 bp (Figure 3.3). Based on this bimodal distribution, we divided the RDs into two groups, one where the donor and acceptor were within 5 kb of each other (local duplications) and the other where the donor and acceptor were separated by > 5 kb or were on different chromosomes (distant duplications). This grouping is further supported by statistical comparison, as the acceptor sizes (mean = 162 bp and 302 bp, max size = 619 bp and 3,424 bp, respectively) were significantly different (Student's t-test, $p < 0.001$).

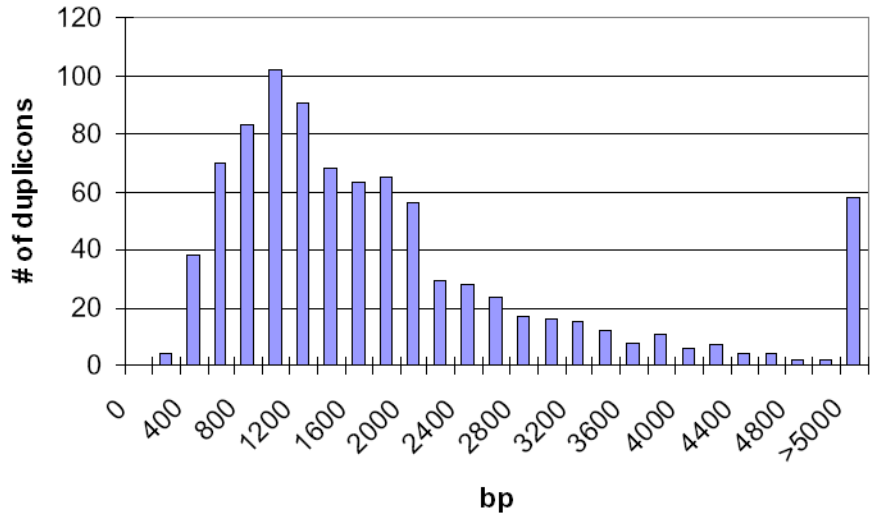


Figure 3.3 Chromosomal distance between donor and acceptor sequences located on the same chromosome in human.

Detailed analysis of the acceptor sequences revealed that in 20 instances, the acceptor was a chimera of 2 different donor sequences, as has been observed experimentally (D'Anjou

et al. 2004), with one donor within 5 kb of the acceptor. In 5 of these cases, the second donor was within 3.5 kb the first donor (range = 289-3,432 bp), while in the other 15 cases, the second donor was located on a different chromosome (Figure 3.4). This type of chimeric acceptor sequence may be formed when the 3' overhanging ends at the site of the DSB invade different template strands and reanneal at a stretch of microhomology.

```
Acceptor Locus:  chr13:106474058-106474295
Donor Locus #1:  chr13:106475923-106476129
Donor Locus #2:  chrY:8097050-8097073

Human:     aaGCAGCCACTTATTATTTG...CTTAAGTTTGTAAGTTTGTA...TAAGTTTGTAtttcctctcg
Chimp:     aagcagccac----------...--------------------...----------tttcctctcg
Rhesus:    aagcagccac----------...--------------------...----------tttgctcccg
Consensus: AACTGGCTGN----------...--------------------...----------TWTCCGC-CG
Donor #1:            TTATTATTTG...CTTAAGTTTGTAA
Donor #2:                                ATTGTTTGTA... TAAGTT
```

Figure 3.4 RD where acceptor is a chimera of 2 different donor sequences. Pre-insertion empty sites for the human-specific RD are shown for chimp and Rhesus macaque. The sequences for donor #1 and donor #2 are shown at the bottom, with the regions of homology between the 2 donors in bold and italics. The filler sequence is in bold and underlined.

3.2.2 Analysis of Duplication Breakpoints

We next focused on precisely defining the breakpoints of the potential RDs in our dataset to determine if the hallmarks of SDSA repair were present. For this, we took advantage of the availability of draft genome sequences for four closely related primate species (chimp, orangutan, Rhesus macaque and marmoset) and three non-primate mammals (dog, cow and mouse) from the UCSC Genome Browser to perform a comparative genomic analysis. We found that 24 of the duplicons were human-specific (present only in human, absent in all other primate species), 67 were hominin-specific (present only in human and chimp), 190 were hominid-specific (present only in human, chimp and orangutan), 289 were catarhinne-specific (present only in human, chimp, orangutan and Rhesus macaque), 513 were primate-specific (present only in primate species) and the remaining 53 were present in all primates and at least one other non-primate mammal.

Given the relatively recent divergence of hominids from the primate ancestor (18 myr), and thus, the short time period for substitutions to accumulate, we focused on precisely defining

the breakpoints for both human- and hominid-specific duplications. The breakpoints of duplicons for which orthologous sequences could be confidently identified between species were manually inspected. This included 18 of the 24 human-specific duplications (14 with the donor < 5 kb from the acceptor, 4 where the donor was on a different chromosome) and a random sample of 50 hominid-specific duplicons (25 with the donor < 5 kb from the acceptor and 25 where the donor was on a different chromosome). We found that 51 of the 68 breakpoints (75%) examined were characterized by the presence of at least one of the molecular hallmarks of imperfect repair events including deletions, "filler" sequences or stretches of microhomology between the flanking sequences of the donor and acceptor (Table 3.1a) (Gorbunova and Levy 1997; Liang et al. 1998; Richardson and Jasin 2000b). The 17 remaining acceptor loci were characterized by the addition of a polyA or polyT tract of 5 bp or greater at one end of the insertion, indicative of the retrotransposition of processed mRNA by target-primed reverse transcription (Luan et al. 1993; Luan and Eickbush 1995). Such retrotransposition events are typically accompanied by target site duplications and, indeed, we were able to identify such duplications flanking 12 of the 17 acceptor sites terminating in polyA/T tracts (Table 3.1b). We noticed that none of the 39 acceptor sequences from which the donor was located within 5 kb of the acceptor possessed polyA/T tracts (14 human-specific, 25 hominid-specific), while 17 of the 29 in the other group did (59%) (Table 3.1a). Thus, breakpoints of local duplications (where the donor is < 5 kb from the acceptor) bear only characteristics consistent with imperfect DSB repair. Conversely, the majority of the distant duplications possess characteristics of retrotransposition events. Based upon these data, only the duplications where the donor was < 5 kb from the acceptor could be confidently classified as RDs. Therefore, only this group was considered for further analysis, leaving a dataset of 824 probable RDs.

Table 3.1 Summary of RD breakpoint characteristics. (a) Breakpoints without polyA tails. (b) Breakpoints with polyA tails. (c) Size ranges in bp of each type of characteristic. MH = microhomology, Del = deletion, Filler = filler sequence, TSD = target site duplication, polyA = polyA/polyT tail of 5 bp or more.

	Characteristics	# where distance between donor and acceptor < 5 kb	# on different chromosome
a.	MH, Del	9	1
	MH, Filler	5	1
	MH, Del, Filler	10	0
	MH, TSD	3	0
	MH, TSD, Filler	2	1
	MH only	2	0
	Filler, Del	5	1
	Filler, TSD	0	5
	Filler, Del, TSD	1	0
	Filler only	2	2
	TSD only	0	1
	Total	**39**	**12**
b.	polyA, MH	0	1
	polyA, TSD, Filler	0	2
	polyA, Filler	0	0
	polyA, TSD	0	8
	polyA only	0	2
	polyA, MH, Del	0	1
	polyA, Del	0	1
	polyA, TSD, MH, Filler	0	0
	polyA, TSD, MH	0	2
	Total	**0**	**17**
c.	**Size Ranges (bp)**		
	Del	1-15	
	Filler	1-68	
	MH	2-9	
	TSD	4-21	
	polyA	> 5 bp	

3.2.3 Rate and Timing of Repair-mediated Duplication during Primate Evolution

In order to estimate the rate of repair-mediated duplication during primate evolution, we used the same computational pipeline described above to uncover chimp-, orangutan- and Rhesus-specific RDs. For these analyses, as before, we considered only those duplications where the donor and acceptor were within 5 kb of each other, since manual inspection of the breakpoints of chimp-specific acceptor sites where the donor and acceptor were on different chromosomes yielded similar results to those obtained for human (data not shown). From these data, we were able to infer the number of RDs that have occurred, and thus the rate of repair-mediated duplication, along different evolutionary branches of the primate phylogenetic tree over the past 30 myr (Figure 3.5). In the 12 myr separating the divergence of Rhesus macaque from the hominid lineage, the rate was found to be 11.8/myr (142 hominid-specific RDs). However, the rate decreased along each branch of the hominid clade. In the 12 myr between the divergence of orangutan from human and chimp, the rate was only 3.1/myr. The rates in the three hominid species were almost identical (human = 2.5/myr, chimp = 4.0/myr, orangutan = 2.7/myr), while a significantly higher rate was found in the Rhesus macaque lineage (8.1/myr). These data indicate that there was a substantial slowdown (about 4 fold) in the rate of repair-mediated duplication in the hominid lineage as compared to the period predating the split of the hominid lineage from Old World monkeys.

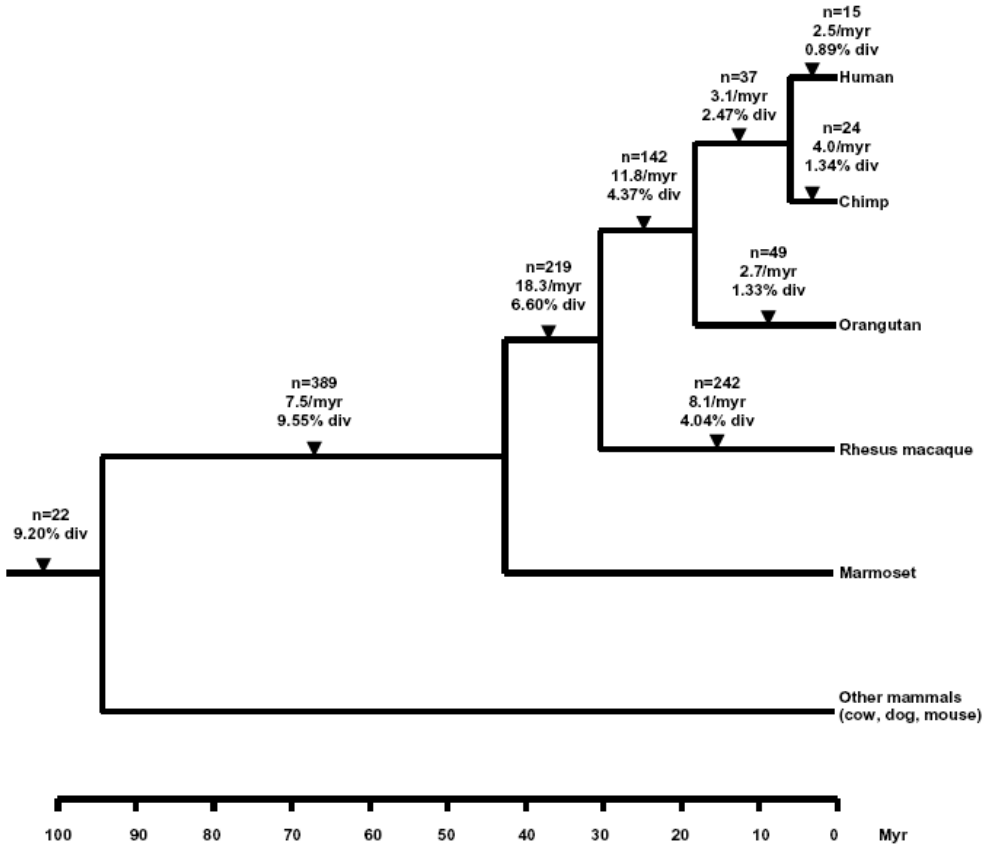


Figure 3.5 Rate and timing of RD formation. Above each branch, the number of RDs created (n), the rate of repair-mediated duplication (per myr) and the average divergence of RDs (%) during the time period are shown.

A good correlation was observed between the average sequence divergence between donors and acceptors and the time period in which the duplication occurred, as inferred from comparative genomics. As expected, the average divergence between donor and acceptor increased as the time since RD formation extended back from the present. The divergence times were also in good agreement with the known sequence divergence between the related species. For example, the average pairwise divergence between donors and acceptors was 0.89% for human-specific and 1.34% for chimp-specific RDs (Figure 3.5). Assuming neutral evolution of RDs and a neutral substitution rate of 2.2x10-9 substitutions/yr (Hedges and Kumar 2004), the expected average divergence for RDs in the human and chimp lineages should be 1.32%. While the average divergence for chimp is almost exactly what is expected, the average

divergence is slightly lower than expected for the human-specific RDs. However, the lower than expected divergence in human may be the result of a small sample size (n=15) and therefore must be evaluated with caution.

Interestingly, the average percent divergence of RDs in the orangutan lineage (1.33%) is significantly lower than expected (3.96%) assuming the same neutral substitution rate as in the human and chimp lineages ($p < 0.05$, $\chi 2$ test). This low percent divergence may be the result of a general slowdown of the neutral substitution rate in the orangutan lineage or of a period of relative quiescence of the mechanism responsible for RD formation early in the orangutan lineage followed by a subsequent increase.

3.2.4 RD Polymorphism and Human Genetic Variation

Identification of human-specific RDs in which the donor and acceptor sequences were identical suggested that these duplications may have occurred in the recent past and could still be polymorphic in the human population. To test this hypothesis, we screened for the presence/absence of 10 human-specific acceptors in 80 individuals from 4 different human geographic populations (African-American, Asian, European and South American) using PCR with primers flanking the insertion. The other 5 acceptors were not tested because specific primers could not be designed. Of these 10 RDs, all but one appeared to be fixed in the human population. Interestingly, one acceptor, present at chr15:31,987,740-31,988,005 in the 2006 hg18 assembly of the human genome, is apparently fixed in all European and South American populations, but remains polymorphic in African-American and Asian populations (Figure 3.6a-d). This acceptor sequence was also precisely absent from the Celera genome assembly, while all other human-specific RDs were present (data not shown). In addition, all 10 RDs examined were absent in the chimp, gorilla and orangutan DNA analyzed (example Figure 3.6e), corroborating our computational prediction that they were indeed human-specific.

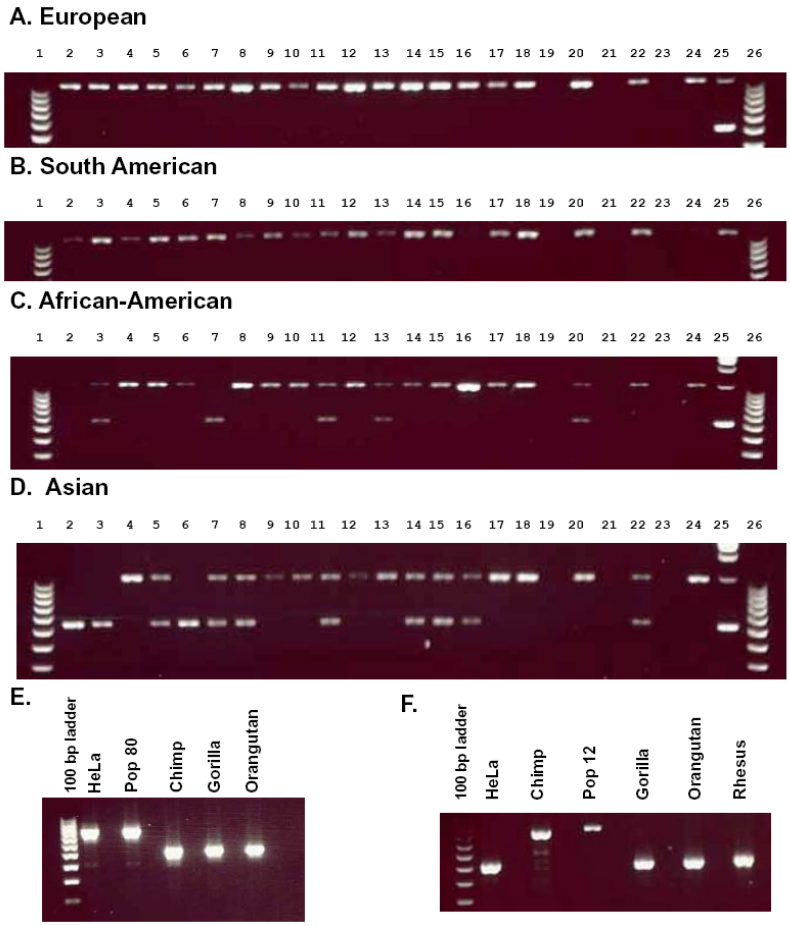


Figure 3.6 (a-d) Experimental verification of a polymorphic human RD at chr15:31,987,740-31,988,005 in the 2006 hg18 assembly of the human genome. For all 20 European (a) and 20 South American (b) individuals surveyed, fragments of expected sizes were obtained for the presence of the acceptor sequence, indicating that all are homozygous for the presence of the acceptor. However, in 20 African-American (c) and 20 Asian individuals (d) surveyed, fragments of expected size for the presence of the acceptor were obtained in some individuals, indicating they are homozygous for the presence of the acceptor (lanes 4-6 in (c) and 4, 9 and 10 in (d)). Fragments of expected size when the acceptor was not present were also obtained for some individuals, indicating they are homozygous for the absence of the acceptor (lane 7 in (c) and 6 in (d)). In addition, fragments of both sizes were obtained in some individuals, indicating they are heterozygous for the presence of the acceptor (lanes 3, 11 and 13 in (c) and 5, 7 and 8 in (d)). See Table 3.4 in Methods for full loading orders. (e-f) Verification of RD lineage specificity. (e) Fragments of expected sized were obtained in human cells (HeLa and Pop80), indicating the presence of the acceptor sequence, while shorter fragments were obtained in chimp, gorilla and orangutan, indicating the absence of the acceptor. (f) Fragments of expected sized were obtained in chimp cells (chimp and Pop12), indicating the presence of the acceptor sequence, while shorter fragments were obtained in human (HeLa), gorilla and orangutan and Rhesus macaque, indicating the absence of the acceptor.

Next, we screened 6 chimp-specific acceptors for polymorphism with DNA extracted from 12 unrelated common chimpanzees, but were unable to find any polymorphic RDs. These results may be due to a small chimpanzee sample size (12 individuals) in comparison to the human sample size (80 individuals). Therefore, the possibility cannot be excluded that some chimpanzee-specific RDs are still polymorphic. Furthermore, each chimp-specific acceptor tested was clearly absent from human, gorilla, orangutan and Rhesus macaque DNA tested (example Figure 3.6f), once again validating that these RDs are indeed specific to chimpanzees.

3.2.5 Duplication of potentially functional sequences

We next investigated whether RDs were responsible for the duplication of functional sequences such as exons, predicted transcription factor binding sites and mammalian most conserved (PhastCons, 28-way) sequences (Table 3.2). We found that 2 complete exons were duplicated, one within the acyl-CoA synthetase bubblegum family member 2 (ACSBG2) gene (exon 7) and the other in a predicted gene (C17orf57, exon 20). In the ACSBG2 gene, the duplicated exon inserted within an intron of the same gene, but in the opposite orientation as the donor exon. Transcriptome data indicates that this duplicated exon is transcribed in the opposite orientation from the donor and appears to form a fusion transcript with an additional exon located upstream of the ACSBG2 gene boundary (EST #CD687637). Ka/Ks analysis revealed no evidence of selective constraint at the coding level since this duplication occurred (Ka/Ks not significantly different from 1, $p > 0.05$). No evidence for the transcription of the duplicated exon in the C17orf57 gene could be found.

Table 3.2 Functional sequences within acceptor and donor sequences

Type of sequence	# in acceptor	# in donor
Intron	295	278
3' UTR	2	11
5' UTR	0	5
Partial exon/intron	0	4
Full exon	0	2
Intron/Full exon/intron	0	2
Mammalian most conserved sequence (PhastCons)	45	82
Predicted transcription factor binding site	22	21

Of the 824 donor sequences, 22 contained sequences annotated on the UCSC Genome Browser as computationally predicted transcription factor binding sites (TFBS). In 21 instances, the same TFBS was also computationally predicted within the corresponding acceptor. Seven of the duplicated TFBS display 100% nucleotide identity between the donor and acceptor, while the other 14 display an average of 86% identity. These data suggest that repair-mediated duplication represents a possible mechanism for duplicating TFBS, thereby potentially contributing to the creation or expansion of regulatory networks.

Mammalian most conserved (PhastCons) sequences are sequences that are more conserved between mammalian species than would be expected under a neutral model of evolution. As such, these sequences may potentially be functional elements evolving under purifying selection. Of the 824 RD donors, 82 contained a full PhastCons segment. Together, these data demonstrate that the mechanism responsible for the creation of RDs can duplicate potentially functional sequences.

3.2.6 Genomic distribution of RDs

Theoretically, it seems that RDs can be created anywhere a DSB occurs. However, DSBs may not occur with the same probabilitiy everywhere in a genome. Thus, the distribution of RDs may reflect, in part, the genomic distribution of DSBs. We sought to discover if RDs

have a bias for certain genomic locations, such as gene-rich or gene-poor regions. To do this, we first plotted the chromosomal distribution of all RDs (Figure 3.7) using the Ensembl Karyoview program (www.ensembl.org). RDs were found on all human autosomes and the X chromosome and were distributed in the expected proportions according to the percentage of the total genomic DNA accounted for by each chromosome. The only exception was chromosome 8, which showed a slight, but statistically significant enrichment of RDs (χ2 test, $p<0.05$, obs=55, exp=39).

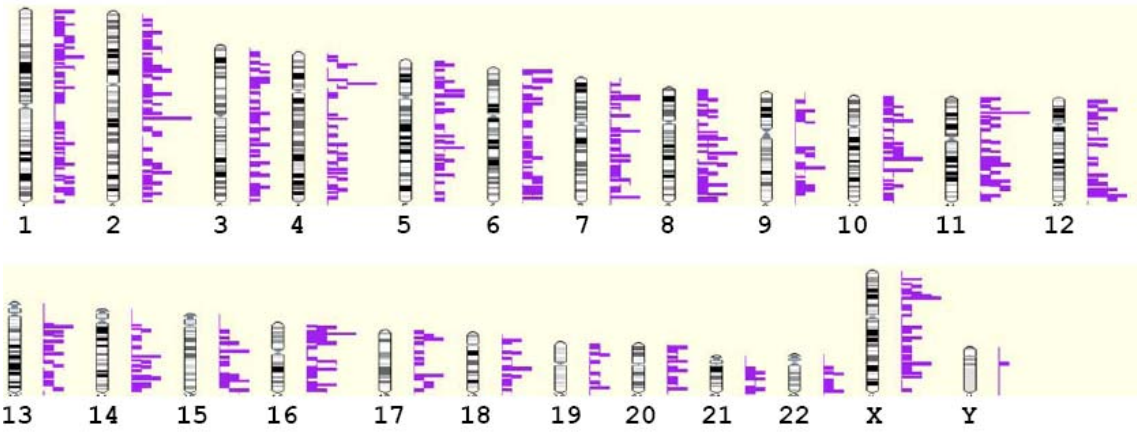


Figure 3.7 Chromosomal distribution of RDs in human. The histogram to the right of each chromosome indicates the number of RDs within the region.

To rule out the possibility that this exception was an artifact of our computational method for finding RDs within TEs, we calculated the total amount of DNA per chromosome occupied by TEs. Using this value to derive the expected number of RDs per chromosome, we obtained the same enrichment on chromosome 8, as well as an enrichment on chromosome 16 (χ2 test, $p<0.05$, obs=32, exp=22). Using this method, a deficiency of RDs was apparent on chromosomes 19 (obs=9, exp=18) and X (obs=29, exp=53). Therefore, while most RDs appear to be randomly distributed in the human genome, two different statistical tests demonstrate an enrichment of RDs on chromosomes 8.

Statistical analysis revealed that RDs occur randomly within the genome with respect to local gene density. For the 824 human RDs, we found no statistically significant difference in

the number of genes located within a 4 MB window (2 MB upstream and downstream) of acceptor sites (mean = 14 per 2 MB, 1 gene per 137 kb) and 5,000 randomly selected sequences with a length of 162 bp, the average length of acceptors (mean = 14 per 2 MB, 1 gene per 138 kb) (Student's t-test, $p > 0.05$). The density of genes per kilobase near RDs is in good agreement with a previous report (Lander et al. 2001) that found an average of 1 gene per 150 kb.

If RDs occur randomly within a genome, then they should be found in TEs in relative proportion to the amount of genomic DNA occupied by each class of TEs. In order to investigate this question, we surveyed 1,000 randomly selected TEs from each major class of TE (SINE, LINE, LTR and DNA). As previously reported (Medstrand et al. 2002; Lander et al. 2001), Alu elements were found in the most gene-dense regions and L1 elements were found in the most gene-poor regions. We found that LTR and DNA elements were located in regions that did not differ significantly from the randomly selected genomic sequences (data not shown). We found a significant paucity of RDs within LINE elements ($\chi 2$ test, $p<0.05$, obs = 293, exp = 377), which are known to accumulate in AT-rich, gene-poor regions, but no deviation from the expectation of the number of RDs found in SINE elements, which, over time, accumulate in gene-rich regions (Medstrand et al. 2002; Cordaux et al. 2006a). RDs were found in significantly higher numbers in LTR (1.4x) and DNA (2.0x) elements than expected (χ 2 test, $p<0.05$).

One possible explanation for the unequal distribution of RDs within the different classes of TEs may lie within Alu elements. Previous reports demonstrate that Alus preferentially insert within AT-rich, gene-poor regions, but accumulate over time in gene-rich regions. Medstrand et al. proposed that Alus within the gene-poor regions are removed over time due to Alu-Alu recombination events (Medstrand et al. 2002). These recombination events in gene-poor regions would most likely have a less detrimental effect than similar events in gene-rich regions. Over time, the Alus within the gene-poor regions are lost, resulting in the observed "shift" of

older Alus to gene-rich regions. These Alu-Alu recombination events in gene-poor regions likely remove surrounding regions where LINE elements are found as well. As a result, RDs within LINEs and SINEs in gene-poor regions could potentially be removed, most likely at a much faster rate than RDs within SINEs located in gene-rich regions. LTR and DNA element mediated recombination events are most likely less frequent than Alu-Alu events since there are fewer copies of these elements to recombine with each other than Alus (Alu: n = ~1,100,000, LTR: n = ~650,000, DNA: n = ~390,000). This lack of recombination between similar LTR or similar DNA elements may result in the accumulation of RDs in these elements due to the lack of a mechanism to remove them. If RDs are selectively neutral, then, like transposable elements, they can stay in the genome for long periods of time without being removed.

3.2.7 RDs in Other Vertebrate Genomes

With strong evidence that RDs are common in primate genomes, we used our computational pipeline to screen 6 additional sequenced vertebrate genomes for the presence of RDs. These genomes included one bird (chicken), one fish (zebrafish) and four other mammal species (mouse, rat, dog and cow). Although the amount and density of TEs in each genome differed significantly among these species (from 8% in chicken to 41% in dog), our pipeline proved effective for uncovering RDs in all genomes surveyed (Table 3.3). For those species where there was sufficient sequence data from a closely related species, multiple alignments were constructed between the surveyed and related species to precisely examine the breakpoints and validate the RDs (Figure 3.8). For those species where sequence data from closely related species was not available, we aligned the TE in which the RD had occurred with the ancestral consensus sequence to identify the pre-integration empty site. Though this method is not as conclusive as aligning related species, it is still effective for predicting the probable breakpoints (Figure 3.8). The analysis of this group of non-primate species shows conclusively that RDs are not only shaping the genomes of primates, but also of other metazoans.

Table 3.3 Number of RDs per species. The right column shows the number of RDs per megabase of transposable elements sequence in the species.

Species	# of RDs	RDs per MB of TEs
Chicken	6	0.08
Chimp	834	0.65
Cow	202	0.20
Dog	192	0.22
Human	824	0.62
Mouse	266	0.26
Orangutan	722	0.55
Rat	179	0.18
Rhesus macaque	650	0.52
Zebrafish	80	0.19

Figure 3.8 Pre-integration empty sites in non-primate vertebrate species. Microhomologies are in bold and italics, filler is in bold and underlined. Deletions are only underlined.

Comparison of the number of RDs within each surveyed vertebrate species revealed a striking trend. The number of RDs per megabase of TE sequence was substantially higher in all primate species (range=0.52-0.65) than in other mammals (range=0.18-0.26) and the difference

was even more pronounced between primates and non-mammalian vertebrates (range=0.08-0.19) (Table 3.3). These data suggest that the mechanism of repair-mediated duplication and/or the probability of RDs becoming fixed within a population differs significantly between different branches of the vertebrate tree.

3.3 Discussion

3.3.1 Validation of experimental method

In this work, we present the first analysis of duplicated DNA segments that bear the hallmarks of repair-mediated duplication of other genomic sequences. While our investigation focused solely upon RDs that occurred within the portion of the genome derived from transposable elements, there is no reason to believe that our results cannot confidently be extrapolated to the rest of the genome.

With the exception of segmental duplications (SDs), low copy number duplications (<10 copies) are typically not annotated in genomes. However, segmental duplications are defined as being a minimum of 1 kb long, so shorter duplications are not analyzed in SD papers. Analysis of low copy number duplications has typically focused on short (<= 100 bp) duplications that have 100% nucleotide identity (Thomas et al. 2004; Messer and Arndt 2007). Therefore, the medium length duplications (50 - 1000 bp) we analyzed belong to a largely overlooked category. One of the challenges of this analysis is the computational time and power involved in identifying and characterizing short duplications. A major advantage of our method is that it significantly reduces both the time and computational resources needed to find such duplications. In addition, the method provides immediate identification of the template (donor) and duplicated copy (acceptor).

Our analysis is unique in that we combine several different methods for identification, dating and lineage-specific verification of RDs. Identification of nested insertions of TEs within TEs has proven to be an effective method for determining periods of transpositional activity (Pace and Feschotte 2007). By using this proven method in a different manner, we have also

demonstrated that the method is robust and powerful, since it is also applicable to finding duplications not caused by TE activity. In addition, analysis of nested insertions into TEs has been found to be effective for discovery of other genomic phenomenon such as selection for LINE elements on the human X chromosome (Abrusan et al. 2008).

The use of comparative genomics provides a second proven method to not only identify RDs, but to verify lineage specificity (Pace and Feschotte 2007; Pace et al. 2008). Duplications are often found by aligning entire genomes and searching for regions that are present in one species but absent from the other. This method is time consuming and cumbersome since whole genome alignments must be performed, followed by subsequent manual analysis. Our method quickly identifies both the donor and acceptor sequences within any genome using Blastn or Blat (Kent 2002) queries and allows for the subsequent identification of the presence or absence of the acceptor sequence at the orthologous locus within closely related species.

Our work is also unique in that we have used experimentally verified hallmarks of imperfect DNA repair to search for evidence of such repair in sequenced genomes. This reverse method capitalizes upon previous experimental investigations in order to find predicted repair events within any genome.

3.3.2 Genomic Impact of RDs

The 824 RDs we were able to identify in the human genome have duplicated a minimum of 133 kb, and potentially 296 kb (when extrapolated to the entire genome rather than just the portion occupied by TEs), of DNA in the human genome. These duplications include untranslated regions (UTRs), exons, splice acceptor and donor sites, predicted transcription factor binding sites (TFBS) and most conserved (PhastCons) elements (Table 3.2). Each of these RDs has the potential to serve as a functional sequence at the acceptor site. In addition, the duplication of TFBS can lead to changes in chromatin structure at the acceptor site, generation of new genes by duplication of promoter sequences, creation of alternative

transcripts (as with ACSBG2) and the expansion of gene regulatory networks. As a result, insertions at the acceptor sites have a tremendous potential for altering genome architecture.

3.3.3 General Slowdown in RD formation during the primate radiation

Our data demonstrate there was a significant slowdown of repair-mediated duplication during the primate radiation (Figure 3.5). The highest rate of duplication (18.3/myr) appears during the 12 myr between the split of New World monkeys (marmoset) from Old World Monkeys (30-42 mya). Following this time period, there is a slowdown in the rate of RD formation in all primate lineages, decreasing to a minimum of 2.5/myr in the human lineage.

Two different hypotheses may explain this slowdown. First, LINE1 has been shown to be a potent source of DSBs. During the early part of the primate radiation (42-77 mya), there was an explosion of LINE1 activity that generated over 280,000 AluJ elements. In the following period (30-42 mya), LINE1 generated over 330,000 AluSx elements. Since that time, there has been a steady decline in the activity of LINE1. The remarkable activity of LINE1 during these time periods would have created hundreds of thousands potential opportunities for DSBs to be repaired via the SDSA pathway, thereby creating RDs. As LINE1 activity decreased, the potential for repair-mediated duplication to occur would have decreased as well.

A second hypothesis may also explain the high rate of repair-mediated duplication observed between 30 and 42 myr ago. Previous studies have shown that a population bottleneck occurred in primates at the Eocene-Oligocene boundary approximately 35 myr ago due to massive mammalian extinctions (Kohler and Moya-Sola 1999). This population bottleneck, coupled with the high level of LINE1 activity, would have provided a suitable environment not only for RDs to form, but also to reach fixation within the various primate populations.

3.3.4 Genome Size and Number of RDs

Interestingly, we found no statistical correlation between genome size and the number of RDs. A comparison of the number of RDs per megabase of TE sequence found in each

surveyed genome with that found in human revealed a striking paucity of RDs in all species except other primates (see Table 3.3). This is especially puzzling when considering species such as dog and cow where the amount of genomic DNA occupied by the major types of TEs is very similar to human. This finding may be the result of the quality of TE annotation available for non-primate species. If TE annotation in non-human species is not as well curated as that in human, then our computational method may not find certain RDs.

A second explanation is that there are evolutionary differences in the cellular machinery responsible for RD formation. Previous studies have shown that proteins involved in non-homologous end joining are under positive selection in yeast, potentially due to pressures exerted by retrotransposon activity (Sawyer and Malik 2006). Additionally, NHEJ proteins such as Cernunnos-XLF have also been shown to be under positive selection within the human lineage (Pavlicek and Jurka 2006). Since DSB repair proteins in each species would be subject to different selective pressures imposed by not only retrotransposons but also retroviruses and differing environmental conditions, it is reasonable to assume that the mechanism responsible for the creation of RDs has also evolved differently within each species lineage. Therefore, the possibility cannot be excluded that the difference in the number of RDs observed within each species is due to differing selective pressures. In addition, at least within the primate lineage, the high number of RDs may be the result of some selective advantage now realized due to the RD formation process.

A third hypothesis for this dramatically higher number of RDs in primate genomes is that there is increased genomic instability in primates. Studies have shown that both the human and chimpanzee genomes have undergone massive explosions of lineage- specific segmental duplications (SD) in the recent past, with as many as 33% of the human SDs having occurred solely in the human lineage (Cheng et al. 2005). In addition, nuclear DNA sequences of mitochondrial origin (NUMTs) have been a potent source of nuclear genome evolution in primates, with Hazkani-Covo and Graur finding 391 NUMTs at orthologous positions in human

and chimpanzee (Hazkani-Covo and Graur 2007). Ricchetti et al. found 27 human-specific NUMTs, 6 of which are polymorphic in the human population, indicating that NUMTs are still shaping human genome architecture (Ricchetti et al. 2004). These studies illustrate the highly dynamic, and thus unstable, nature of the primate nuclear genome.

3.3.5 RDs and Segmental Duplications

It has been proposed that a mechanism similar to SDSA may be responsible for the formation of segmental duplications (Fiston-Lavier et al. 2007). However, we found no evidence that the mechanism responsible for RD formation is the same as that for segmental duplications and thus believe that SDs and RDs are mechanistically unrelated. Three lines of evidence support this claim. First, the largest unequivocal RD we could find was 619 bp with the average size of the acceptor sequences being 162 bp. While this mechanism may create duplications that could be classified as SDs, these events would seem to be atypical. Second, in a sample of SDs we analyzed, all possessed polyA tails and appeared to be the result of retrotransposition (data not shown). Third, the chromosomal distribution of RDs is different than segmental duplications. SDs have been found to be preferentially located at pericentromeric regions with the majority of the duplications located on different chromosomes. RDs are found randomly distributed along each chromosome, with no bias for a particular region, and a strong bias for the duplications being closely linked (Figure 3.7) (for review, see (Bailey and Eichler 2006)).

Our findings that RDs occur not only in primate genomes, but also in other vertebrate genomes, indicates that this mechanism has been shaping genomes for potentially hundreds of millions of years. Since the RDs are putatively created by SDSA, a combination of homologous recombination and non-homologous end joining, both of which have been found to occur in a wide range of eukaryotic genomes, it is not surprising that this mechanism is so widespread. Thus, classification of RDs as a major force in eukaryotic genome evolution is not

unreasonable. Further investigation of RDs within newly sequenced genomes is therefore warranted.

3.4 Methods

3.4.1 Retrieval of Genome Sequences and RepeatMasker rmsk files

Genome sequences and RepeatMasker rmsk files were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu). The versions used were: human (hg18), chimp (panTro2), Rhesus macaque (rheMac2), mouse (mm8), rat (rn4), dog (canFam2), cow (bosTau2), chicken (galGal3) and zebrafish (danRer4).

3.4.2 Identification of RDs

Potential RDs were identified by a Perl script that searched the RepeatMasker rmsk files for TEs that had been interrupted by some intervening sequence. A TE was classified as interrupted if the repeat name and orientation of the first segment (TE-A) matched the repeat name and orientation of the second segment (TE-B), TE-A and TE-B were separated by at least 50 bp and neither TE-A nor TE-B was longer than 95% of the length of the consensus sequence. In addition, the ending consensus sequence position of TE-A was within +/- 30 bp of the starting consensus sequence position of TE-B. After the potential RDs were identified, false positives were removed. TEs separated due to nested TE insertions were removed with a Perl script and annotated retrogenes, segmental duplications or LINE-1 3' transduction events were manually inspected and removed from the dataset.

For all remaining potential RDs, a Perl script retrieved the acceptor sequence, along with 100 bp flanking each side, and used this sequence as a Blastn query against the entire genome. In order for a Blast hit to be considered, it had to match at least 80% of the length of the query sequence with at least 50% identity. The minimum cutoff score was calculated separately for each acceptor sequence using a sliding e-value that was unique for each query sequence. The Blast output was then parsed. If the acceptor sequence minus the flanking regions was found in more than two HSPs, the RD was removed from the dataset since the

donor could not be unequivocally determined. Additionally, if the acceptor sequence with the flanking regions was found more than once, the RD was discarded to avoid classifying potential segmental duplications or transposition of chimeric TEs. Finally, if the donor was within 50 bp of the acceptor sequence, the RD was also removed from the dataset.

If all of the above criteria were successfully met, the putative donor sequence was used as a Blastn query sequence against the entire genome. This "reciprocal" Blast query then used the same criteria to determine if any sequence matched the donor. In order for the acceptor and donor sequences to be classified as a putative RD, the acceptor sequence had to be the second hit in the Blastn output generated when the donor was the query sequence and the acceptor had to meet all criteria.

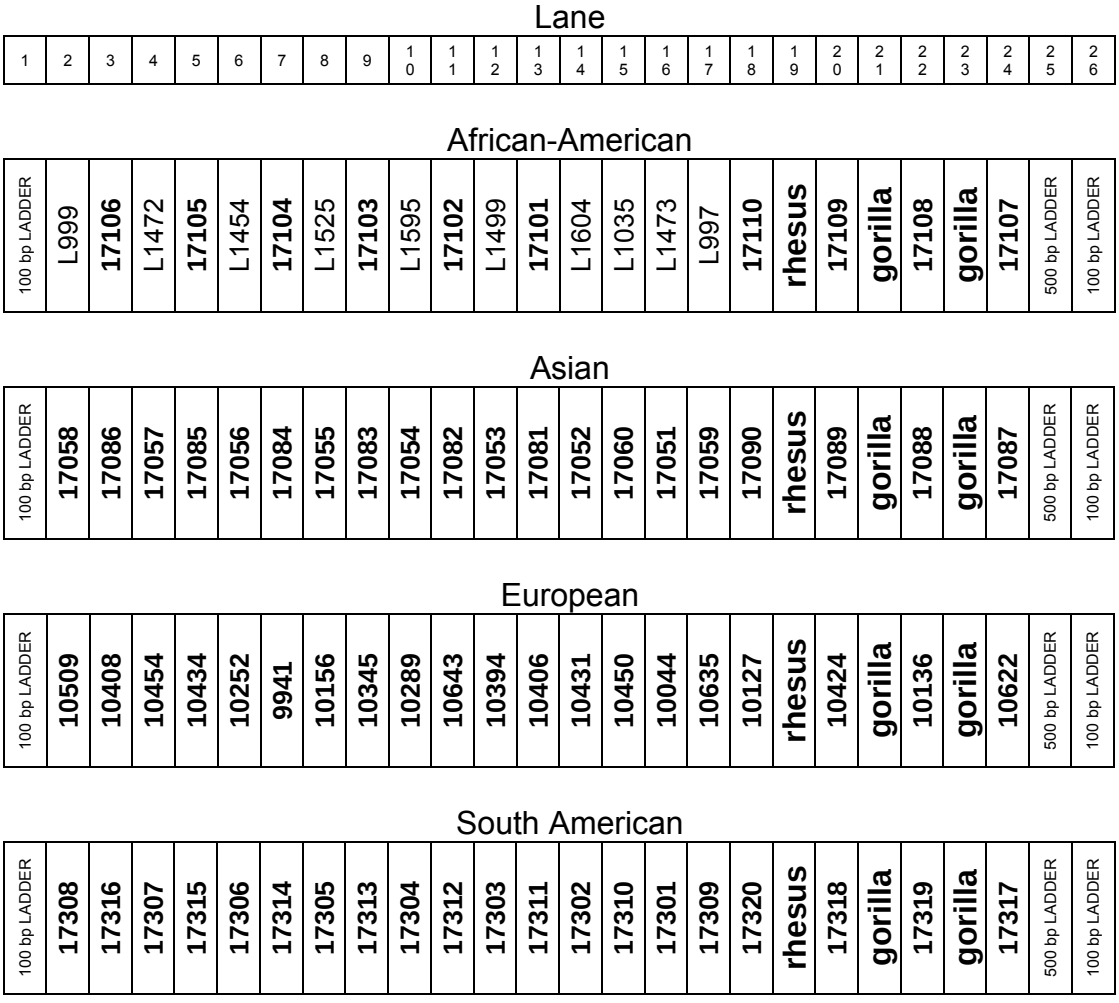
3.4.3 PCR Verification of Computationally Detected RD Loci

To verify that the computationally detected RDs existed in vivo and did not represent genome assembly errors, we designed oligonucleotide primers flanking each locus using the Primer3 web interface (http://frodo.wi.mit.edu/). PCR amplification was performed in 25-ml reactions with 10–50 ng genomic DNA, 200 nM of each oligonucleotide primer, 200 mM dNTPs in 50 mM KCl, 1.5 mM MgCl2, 10 mM Tris-HCl (pH 8.4), and 2.5 units Taq DNA polymerase on an Applied Biosystems GeneAmp® PCR System 9700 thermocycler. Amplification cycles were as follows: an initial denaturation step of 94ºC for 4 min; followed by 32 cycles of 1 min of denaturation at 94ºC, 1 min of annealing at optimal annealing temperature, and 1 min of extension at 72ºC; followed by a final extension step at 72ºC for 10 min. For loci with large duplications (>2 kb), we used Ex Taq™ polymerase (TaKaRa) and carried out PCR in 50 ml reactions following the manufacturer's suggested protocol. PCR amplicons were separated on 2% agarose gels, stained with ethidium bromide, and visualized using UV fluorescence.

To identify lineage-specific human and chimpanzee duplication loci, PCR amplification were performed on a panel of five primate species, including *Homo sapiens* (HeLa; cell line ATCC CCL-2), *Pan troglodytes* (common chimpanzee; cell line AG06939B), *Pan paniscus*

(bonobo or pygmy chimpanzee; cell line AG05253B), *Gorilla gorilla* (western lowland gorilla; cell line AG05251), and *Pongo pygmaeus* (orangutan; cell line ATCC CR6301). To evaluate polymorphism rates of human lineage-specific duplications, we amplified loci on a panel of genomic DNA from 80 human individuals (20 from each of four populations: African-American, South American, European, and Asian) that was available from previous studies in the Batzer lab at Louisiana State University (Table 3.4).

Table 3.4 Loading order of gels



CHAPTER 4

WIDESPREAD HORIZONTAL TRANSFER OF DNA TRANSPOSONS IN MAMMALS

4.1 Introduction

Transposable elements are segments of DNA that can insert into different locations in a genome, often duplicating themselves in the process. Class I, or RNA-intermediate, retrotransposons transpose through a copy-and-paste mechanism that always results in the addition of a new copy of the element. Conversely, Class II, or DNA-intermediate, transposons move through a cut-and-paste mechanism. This method of transposition may result in the duplication of the element or simply the movement of the element from one genomic location to another, depending on how the lesion where the transposon was excised is repaired.

TEs have been characterized as “selfish” and “parasitic” genetic elements (Orgel and Crick 1980) that confer no selective advantage to their host and therefore evolve as “neutral genomic residents” (Cordaux et al. 2006a). Since TEs are not typically under selection to be maintained in the genome, they are subject to the accumulation of substitutions which eventually leads to their death, or inability to transpose. However, in certain instances, all or part of a TE may become domesticated. In these cases, the element now serves a beneficial role and is no longer evolving neutrally, but is subject to selection (Cordaux et al. 2006b; Kapitonov and Jurka 2005). In order to avoid extinction, TEs differentially utilize two distinct mechanisms to ensure their survival, or continued ability to transpose: vertical diversification and horizontal transfer.

Phylogenetic analyses in widespread eukaryotic species indicate that the first type of Class I elements, the non-LTR retrotransposons, typically survive through vertical diversification (Malik et al. 1999; Gentles et al. 2007; Khan et al. 2006), for review see (Furano 2000), with one

element family evolving into a new family, eventually supplanting its predecessor (Cabot et al. 1997; Saxton and Martin 1998). However, evidence for the horizontal transfer (HT), or movement of DNA between reproductively isolated species, of the RTE non-LTR retrotransposon between ruminants and snakes has been demonstrated (Zupunski et al. 2001; Kordis and Gubensek 1998) and preliminary data suggests that the RTE elements in opossum may have been horizontally transferred as well (Gentles et al. 2007). The mechanism for the HT of non-LTR retrotransposons remains unclear, but Piskurek and Okada (2007) identified a SINE element that was transferred into a poxvirus by retrotransposition, allowing the virus to serve as the vector to transfer the retroposon to mammalian species.

The second type of Class I elements are the LTR retrotransposons. These elements are similar in structure to retroviruses, possessing ORFs for both gag and pol genes, yet lacking a third ORF for an envelope (env) gene. If the element acquires an env gene from an exogenous source, its horizontal transmission can occur via infection (Malik et al. 2000). Song et al. showed that when a strain of Drosophila larvae possessing no active gypsy elements, an LTR retrotransposon that has acquired an env gene, are exposed to gypsy virus particles, high levels of insertion are observed (Song et al. 1994). However, other studies have demonstrated the HT of the copia (Jordan et al. 1999; Jordan and McDonald 1998) family of elements in Drosophila as well as the SURL family in sea urchins (Gonzalez and Lessios 1999), neither of which possess an ORF for an env gene, via an unknown mechanism that does not involve infectious viral-like particles. Therefore, the data strongly suggests that the survival of LTR retrotransposons may be achieved through 2 separate forms of horizontal transfer: enzymatic machinery encoded by the element that allows them to become infective virus particles or through an unknown vector, potentially an exogenous RNA or DNA virus (Friesen and Nissen 1990).

The survival of DNA transposons is hypothesized to rely heavily upon horizontal transfer into virgin genomes (Lander et al. 2001). Indeed, the HT of DNA transposons in

eukaryotes has been well-documented, particularly in insects (for reviews see (Silva et al. 2004; Robertson 2002; Hartl et al. 1997; Loreto et al. 2008). In Drosophila, HT of the P (Clark and Kidwell 1997; Clark et al. 1994; Haring et al. 2000; Daniels et al. 1990), *mariner* (Maruyama and Hartl 1991) and *hobo* (Simmons 1992) superfamily elements has been demonstrated. Additionally, mariner elements have been shown to be horizontally transferred between insects of differing orders (Robertson and Lampe 1995; Robertson and MacLeod 1993). One example of a Mutator superfamily has been shown to have been horizontally transferred between plant species (Diao et al. 2006). Two cases of the HT of DNA transposons into vertebrate species have been documented, with both instances occurring in fish (Koga et al. 2000; de Boer et al. 2007). With the discovery of the *SPIN* element and its broad taxonomic distribution between not only diverse mammals, but also distantly related tetrapod species, it became apparent that DNA TEs have been capable of invading multiple mammalian genomes during the mammalian radiation (Pace et al. 2008).

To investigate if the horizontal transfer of DNA transposons into diverse mammalian species is a recurrent phenomenon, we sought to identify TEs that, like *SPIN*, had invaded multiple mammalian species. We discovered that six known mammalian DNA transposon families (Charlie12, Charlie3, HSMAR1, HSMAR2, OposCharlie1 and RCHARR1), previously described as lineage-specific (Pace and Feschotte 2007; Lander et al. 2001; Robertson and Zumpano 1997; Robertson and Martos 1997; Jurka et al. 2005; Waterston et al. 2002), were actually found in multiple mammalian species. Each displayed (i) a patchy taxonomic distribution, (ii) a level of intra-specific similarity lower than would be expected if the element was present in a common ancestor and was evolving at or near the neutral substitution rate and (iii) a level of inter-specific similarity of the consensus sequences between species that is much greater than expected considering the divergence times of the species. In addition, we identified 4 previously unannotated families of DNA transposons (Dumbo, LAMAR2, OGMAR1 and OGMAR2) that have been horizontally transferred. These results suggest that the HT of

DNA transposons into multiple mammalian species has occurred repeatedly, contributing to bursts of TE activity and independently shaping genome architecture in mammals.

4.2 Results

4.2.1 Discovery of new DNA TE families

We began our investigation by assembling a comprehensive database of all known DNA TEs in 19 mammalian genomes (see Methods). Masking each genome with the database yielded several TEs with patchy distributions. For example, HSMAR2, an autonomous element previously described as primate-specific (Pace and Feschotte 2007), was found in rabbit and pika (Lagomorpha) and tenrec (Afrotheria), but was absent in all other mammalian species. To investigate these findings, consensus sequences for putatively species-specific elements were derived (see Methods). Each new consensus sequence was added to a species-specific database of DNA TEs that included all previous DNA TEs and the new consensus sequence. The genome was then remasked with this database. In addition, the consensus was used as a query sequence in Blastn searches (Altschul et al. 1997) against the other 18 genomes to assess the presence or absence of the element, thereby validating the results obtained from RepeatMasker (Smit et al. 1996-2004).

After masking, the RepeatMasker output was analyzed for additional TEs, typically MITEs, in the same tribe as the newly identified, species-specific element. A tribe is defined as an autonomous family of elements and the corresponding MITE families that were derived from the autonomous element. As new elements were found and their consensus sequences derived, the genome was remasked using the species-specific repeat database as before. This iterative process was done for all species in all cases where previously unidentified elements, or new TEs within a tribe of previously annotated repeats, were found. In total, this process yielded 4 previously unidentified putative autonomous families of DNA transposons (LAMAR2, Dumbo, OGMAR1 and OGMAR2) and their corresponding MITEs in 3 different species (elephant, tenrec and bushbaby). Each of these tribes was found to be specific for one species

or clade. Two TE families, each previously reported as species- or clade-specific, were confirmed as being present in only one particular species or clade (MER75, Oamar1) (Pace and Feschotte 2007; Jurka et al. 2005). Six families of TEs, previously thought to be specific to a single lineage, were found to be present in multiple species with a patchy distribution (Charlie12, Charlie3, HSMAR1, HSMAR2, OposCharlie1 and RCHARR1). The autonomous elements (Charlie12, Charlie3, HSMAR1, HSMAR2 and OposCharlie1) all generated lineage-specific MITEs in at least one species where they were found.

4.2.2 Horizontal transfer of hAT elements

A total of 4 different families of hAT superfamily transposons have been independently horizontally transferred into mammalian species (Charlie12, Charlie3, OposCharlie1 and RCHARR1). Charlie12, Charlie3 and OposCharlie1 are all autonomous elements that spawned distinct MITE families in each genome (Appendix C). RCHARR1 is a non-autonomous element for which the autonomous element responsible for its mobilization could not be identified.

4.2.2.1 Charlie12 family

Charlie12 is the autonomous element that mostly likely gave rise to the human MER30 and MER30b families of elements (Appendix C). All 3 of these families were previously classified as anthropoid-specific based upon their age according to sequence divergence, the age of TEs they inserted into and that inserted into them (nested insertion analysis, or NIA) and cross-species genomic analysis of orthologous loci (CSA). However, masking the other mammalian genomes with the human Charlie12 consensus sequence revealed that Charlie12 is not only present in anthropoid primates but also in murine rodents (mouse and rat), guinea pig, bushbaby and bat, while being clearly absent in all other surveyed species. Analysis of the activity of the Charlie12 family reveals that it invaded at least 3 different mammalian species independently: the common ancestor of prosimian and anthropoid-primates, the ancestor of rodents and the ancestor of bats.

While Charlie12 did not achieve high copy number in any species except bushbaby (n = 3,603, all other species n < 530), it did successfully generate distinct MITEs in each lineage. In total, at least 7 distinct MITE families were identified in the 5 different lineages (MER30 and MER30b in anthropoid primates, MER30_Ml_1, MER30_Ml_2, MER30_Ml_3 in bat, RMER30 in murine rodents and MER30_Og_1 in bushbaby) (Appendix C). Four of these families were previously undescribed (MER30_Ml_1, MER30_Ml_2, MER30_Ml_3 and MER30_Og_1). In all species, at least 1 MITE family was able to attain a copy number of greater than 1,500. In bushbaby, the MER30_Og_1 MITE attained a copy number of greater than 8,100 (Appendix D).

4.2.2.1.1 Charlie12 in primates

In order to discover when Charlie12 invaded the primate lineage, the three methods previously employed to date TEs were used (Pace and Feschotte 2007). If Charlie12 was introduced into the primate lineage before the divergence of tree shrew from the primate ancestor, then we would expect to find at least one copy of Charlie12 or a MITE family within the tree shrew genome. However, no copies could be identified, but pre-insertion empty sites (PIES) in tree shrew into which a Charlie12 family element had inserted in the bushbaby or human genome were recognizable (Figure 4.1). From this we were able to conclude that Charlie12 was introduced into the primate lineage after the divergence of tree shrew, or less than 89 mya.

```
MER30_Og_1 empty site in tree shrew
Bushbaby position:    scaffold_437.1-256026:196373-196981
Tree shrew position:  scaffold_150151.1-150382:13051-13429
Bushbaby:             AAAGCAAAAC GTTTAAAC cagggtgtt...ATGCCCCTG ATTTAAAC CATTCATTTCTTTAAATTGCAC
Tree shrew:           AAAGCAAAAT ATTTAAGC ----------...----------CATT------TTTAA-TTGCAC
```

Figure 4.1 Pre-insertion empty site in tree shrew for bushbaby MER30_Og_1 element. Target site duplications are in bold and underlined.

Two additional possibilities exist for the timing of the invasion of Charlie12 into primates. First, the infiltration could have occurred before the divergence of prosimians from the primate lineage. Second, Charlie12 could have invaded the bushbaby and human lineages independently after the divergence of these species. Evidence points to the insertion of the

Charlie12 family into the common ancestor of prosimian and anthropoid primates, followed by lineage-specific activity.

The first line of evidence is that the MITE families identified in bushbaby (MER30_Og_1) and human (MER30 and MER30b) are structurally different (Appendix E). No orthologous copies of the human MER30 or MER30b elements could be identified in bushbaby. Indeed, no copies of MER30 or MER30b could be found anywhere in the bushbaby genome. However, 2 copies of the bushbaby MER30_Og_1 were found at orthologous loci in human. For MER30_Og_1 to have transposed, at least one copy of Charlie12 must have been present and have spawned at least 2 copies of the MER30_Og_1 element before the divergence, as the probability of the insertion of two elements into the same loci in two different species is negligible. Apart from these 2 copies, no additional MER30_Og_1 elements could be found in human, rather, PIES for MER30_Og_1 could be identified in human and those for MER30 or MER30b in bushbaby (Figure 4.2 a-b). This analysis suggests that Charlie12 infiltrated the genome of the prosimian/anthropoid primate ancestor and continued to be transpositionally active after the divergence of the species, generating lineage-specific activity of MER30_Og_1 in bushbaby and MER30 and MER30b in anthropoids.

```
MER30_Og_1 empty site in human
Bushbaby position:  scaffold_437.1-256026:196373-196981
Human position:     chr6:117818290-117818630
Bushbhaby:          AAAGCAAAAC GTTTAAAC caggggtgtt...ATGCCCCTG ATTTAAAC CATTCATTTCTTTAAATTGCAC
Human:              AAAGTAAAAC ATTTAAAC ----------...----------CATT------TTTAA-TTGTAC
```

(a)

```
MER30 empty site in bushbaby
Human position:     chr21:34098858-34099190
Bushbaby position:  scaffold_106410.1-185642:40384-40822
Human:              ataag catttaaa ccaagcttgtc...agattgtact gatttaga cactc
Bushbaby:           acaac catttaga -----------...----------tgatc
```

(b)

Figure 4.2 Pre-insertion empty sites in (a) human for bushbaby MER30_Og_1 and (b) bushbaby for human MER30 element. Target site duplications are in bold and underlined.

It should be noted that Charlie12 is a low copy number element in all genomes except bushbaby. In human, only 48 highly-fragmented copies are identifiable. None of these fragments are found at orthologous loci in bushbaby. The copy, or copies, of Charlie12 that were responsible for the transposition of MER30_Og_1 in the ancestor of prosimians and anthropoid primates is no longer identifiable between the two species. Because of this, we cannot absolutely say that Charlie12 was present in the prosimian/anthropoid ancestor. However, this is the most likely conclusion given that copies of a Charlie12 MITE are orthologous between the two species.

The second line of evidence comes from the age of Charlie12 and its corresponding MITEs in each lineage. The average age of each family of elements was calculated using the method previously employed to date SPIN elements (Pace et al. 2008) (see Methods). The average ages of Charlie12 (in bushbaby) and MER30_Og_1 were found to be 78 and 66, respectively. The MER30 and MER30b families in human were calculated to be 59 and 54. These dates further support the hypothesis that all of these families had lineage-specific activity, assuming that prosimians and anthropoids diverged 80 mya.

Third, the lineage-specific activity of Charlie12 in primates is corroborated by NIA using Transposon Cluster Finder (TCF), a program that uses nested transposons to determine the age of TEs in relation to each other (Giordano et al. 2007). According to the relative age map of TEs generated by TCF (Figure 4.3), MER30_Og_1 is similar in age to the OGMAR2, SPIN and MER1A_Og families of elements, all of which are prosimian-specific (Appendix D). Indeed, instances of MER30_Og_1 insertions into OGMAR1 (60 myr), OGMAR2 (60 myr) and HSMAR2_Og (65 myr) and Charlie12 into OGMAR2 define the lower bound of the ages and further corroborate the calculated ages of MER30_Og_1 and Charlie12. The upper bound of the age can be determined by searching for the oldest TEs that inserted within a given TE. The oldest element to insert into MER30_Og_1 was Charlie12 and the oldest to insert into Charlie12 was MER30_Og_1. These data allow us to conclude that the activity of Charlie12 most likely

began at approximately 80 mya, or roughly the time of the prosimian/anthropoid divergence, and ended near 60 mya. The combination of the these data support the conclusion that Charlie12 invaded the primate lineage in the time period between the divergence of tree shrew and prosimians, or approximately 80 to 90 mya, and produced lineage-specific activity in both the prosimian and anthropoid primates.

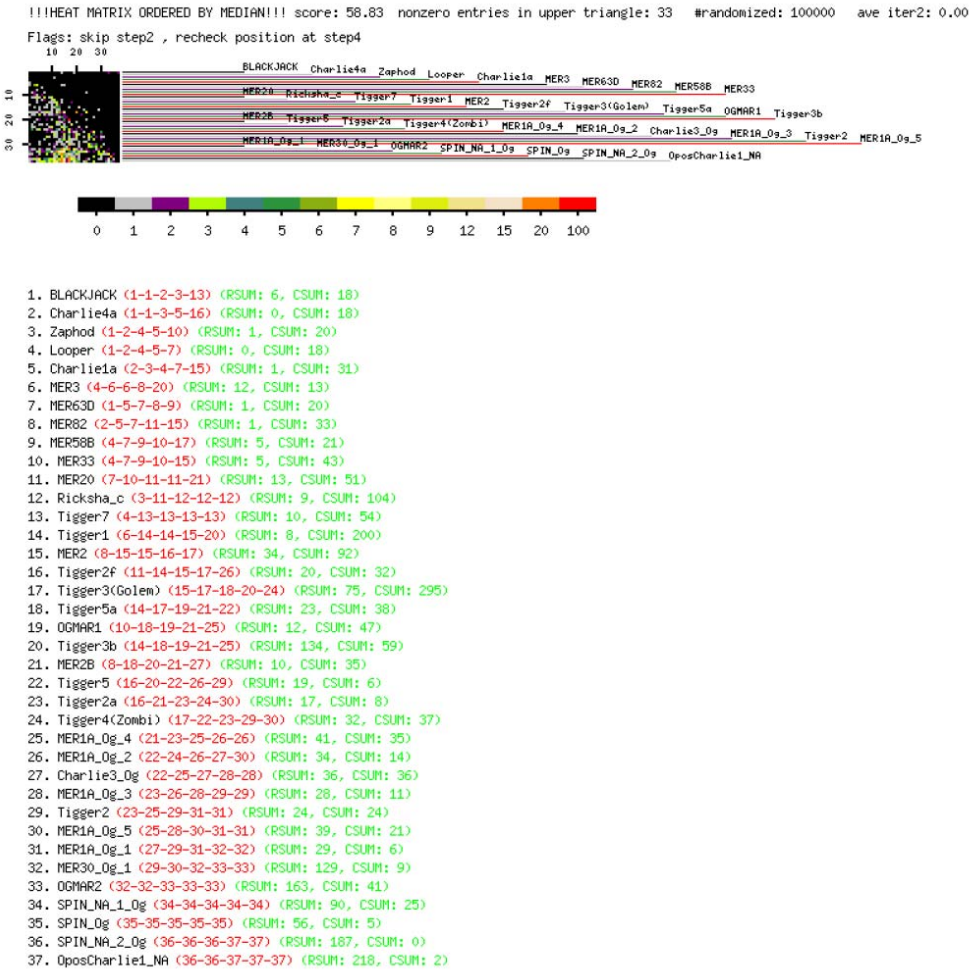


Figure 4.3 Transposon Cluster Finder (TCF) median ordered heat map showing relative ages of DNA TEs in bushbaby using nested insertion analysis.

4.2.2.1.2 Charlie12 in rodents

The same scenario of Charlie12 infiltration prior to the divergence of two species followed by lineage-specific activity may be true in rodents. Both Charlie12 and its rodent-

specific MITE, RMER30, were identified in mouse, rat and guinea pig after masking the genomes with the Charlie12 and RMER30 consensus sequences. CSA of the rodent Charlie12 tribe revealed that over 500 copies of RMER30 are present at orthologous loci in mouse and rat. From this, we can conclude that the element was active prior to the divergence of mouse and rat (15 mya). We can also conclude that RMER30 and Charlie12 are not in the squirrel, rabbit or pika genomes, as masking with these genomes produced no identifiable copies of the elements. We were able to find over 480 copies of RMER30 in guinea pig, which diverged from the rodent lineage ~70 mya. This finding suggests that RMER30 was present in the ancestor of guinea pig and murine rodents. However, we were unable to identify any orthologous copies of RMER30 between guinea pig and mouse or rat.

The lack of identifiable orthologous copies between the species does not necessarily show that the element was not present prior to their divergence for several reasons. First, because the neutral substitution rate in rodents is substantially higher than in other mammals, identification of old TEs and their orthologs in related species is often difficult, if not impossible (Waterston et al. 2002). Thus, the lack of identifiable orthologs may simply be the result of the high substitution rate obscuring our ability to detect orthologs. Second, RMER30 in guinea pig does not appear to be structurally distinct from RMER30 in mouse/rat, unlike MER30_Og_1 and MER30. Once again, this conclusion must be viewed cautiously because reconstruction of a consensus sequence for RMER30 in guinea pig was extremely difficult considering that the average divergence of individual copies is 30.8%. Because of these extenuating factors, the possibility cannot be excluded that RMER30 was present in the rodent ancestor. Conversely, the lack of orthologs between species may indeed be the result of lineage-specific activity and RMER30 in guinea pig may be slightly different than that in mouse. However, because of the high level of uncertainty, a definite conclusion cannot be reached.

NIA of the Charlie12 tribe in both mouse and guinea pig does not provide further support for either lineage-specific or pre-divergence activity of Charlie12. In mouse, the

youngest element that RMER30 inserted into was RMER5 (83 myr), while the oldest TE that inserted into it was B4A (107 myr). Charlie12 also suffered insertions from B4A. In guinea pig, where only DNA elements were masked, no elements inserted into either Charlie12 or RMER30, nor did they insert into any other TEs. Therefore, the bounds for Charlie12 family elements in mouse according to NIA range from ~80 to 110 myr.

The calculated age of RMER30 in mouse is 91 myr and in guinea pig is 94 myr, while the age of Charlie12 is 95 and 96, respectively. These ages are significantly higher than those of the Charlie12 tribe of elements in primates (range = 54 - 78). It is possible that the elements in rodents are much older than in primates, but this seems unlikely. If the rodent elements are truly this old, then we would expect to find them in squirrel and lagomorphs, but we do not. While their absence may be explained by lineage-specific loss or lineage sorting, we believe another explanation is more plausible. The substitution rate used to calculate ages in the rodent lineage (as in all others lineages except anthropoid primates), assumes a constant molecular clock since the divergence of human from rodents. Our previous work has shown that the clock was highly variable during different time periods in the primate lineage (Pace and Feschotte 2007). There is no reason to believe that the same is not true in the rodent lineage as well. However, until more species of rodents are sequenced to provide better resolution, we cannot obtain more accurate calibration of the molecular clock in rodents during different evolutionary periods. Because of this, the ages of older TEs in rodents may be overestimated. If this is true, and the rodent Charlie12 tribe elements are younger than our calculated ages, it would explain why the elements are only found in the guinea pig and mouse lineages, while being absent from squirrel and lagomorphs.

The question of whether there was lineage-specific activity of the Charlie12 tribe in rodents can be further addressed by exploring the evolutionary dynamics of these elements in the two lineages. In mouse, 2,851 copies of RMER30 can be identified, but only 483 in guinea pig. If all Charlie12 tribe activity occurred before the divergence of these species, then we must

conclude that over 2,000 copies of RMER30 were purged from the guinea pig genome. This scenario seems unlikely, especially considering how efficiently the mouse genome removes transposable elements (Waterston et al. 2002). A more probable hypothesis is that the RMER30 MITE was found in the common ancestor and was differentially multiplied in both lineages, with almost 6 times as many copies being created in the murine lineage. This independent amplification could also explain the lack of orthologous copies of RMER30 between the species.

4.2.2.1.3 Charlie12 in bats

The average age of Charlie12 and its respective MITEs in bat (MER30_Ml_1 to MER30_Ml_3) appear to be significantly younger than the MITEs in either the rodent or primate lineages based upon calculated age (range = 29-42 myr) (Figure 4.4). These age calculations are further validated by CSA. First, no Charlie12 elements were masked by RepeatMasker in any other Laurasiatherian species. Second, PIES for MER30_Ml_x elements could be identified in dog (Figure 4.5). Further, NIA ranked MER30_Ml_1 the fifth youngest element, just after the 4 bat SPIN MITEs, all of which have been determined to be less than 21 myr old. MER30_Ml_1 was also found to be relatively younger than MER2 (75 myr) and Tigger1a_Mars (65 myr). These three methods converge to demonstrate that the Charlie12 tribe of elements in bat are indeed between approximately 29 and 42 myr old, and thus, specific to the bat lineage.

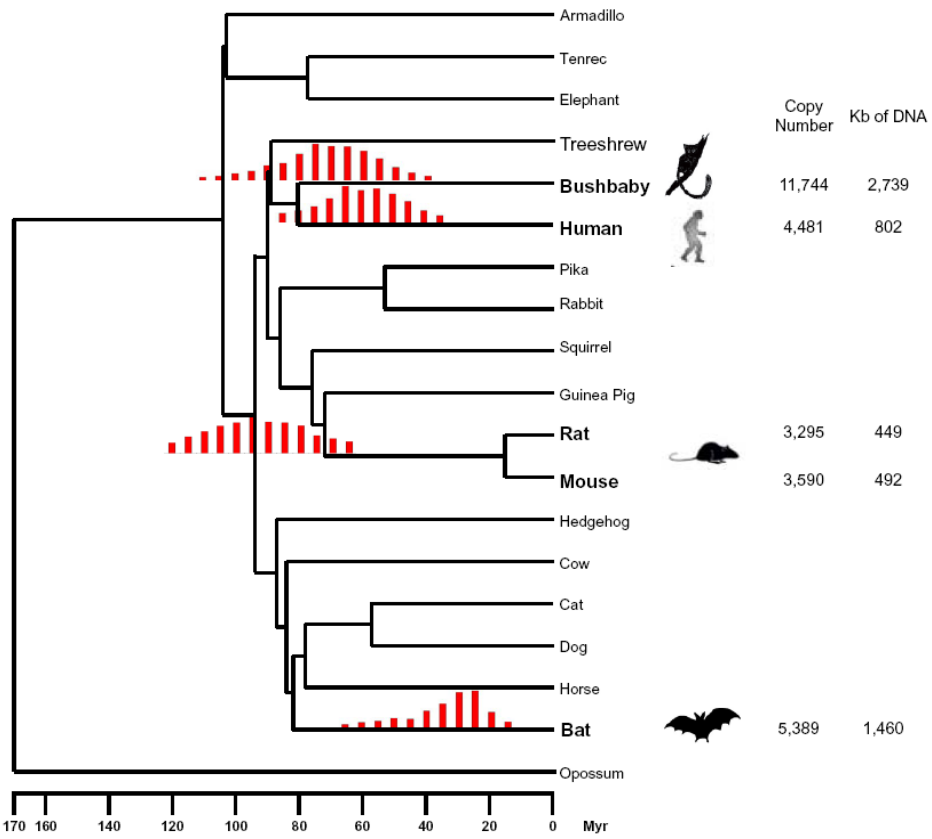


Figure 4.4 Species distribution and timing of amplification of Charlie12 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring Charlie12 tribe transposons are in bold. The timing of Charlie12 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all Charlie12 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

```
MER30_Ml_1 empty site in dog
Bat position:  gi105852675gbAAPE01128715.1:1720-2199
Dog position:  chr10:51515694-51516030
Bat:           TGCATTAATTACCTAGATCAAGCATGTC...GACATGCTTGACCTAGATCTTCAACATC
Dog:           TACATGTGTTGCCTTGAT----------...----------CCCGAACATG
```

Figure 4.5 Pre-insertion empty site in dog for bat MER30_Ml_1 element. Target site duplications are in bold and underlined.

CSA, NIA and dating of elements based on percent divergence clearly show lineage-specific activity of Charlie12 tribe elements in bushbaby, anthropoid primates, rodents and bat. These data highly suggests that the infiltration of Charlie12 into these lineages was independent. In addition, CSA reveals that no copies of any Charlie12 tribe elements from any lineage are found at orthologous loci in any other species where Charlie12 is found. Therefore, it is reasonable to conclude that the Charlie12 family of elements independently invaded the lineages of 3 different clades: primates, rodents and Laurasiatherians. No evidence exists to suggest that Charlie12 is in these lineages as a result of vertical transmission instead of horizontal transfer. In all cases, Charlie12 formed different tribes, producing lineage-specific MITE families and subsequent insertions. In total, these elements contributed 2,739 Kb of DNA to the prosimian lineage, 802 Kb to the anthropoid lineage, at least 449 Kb to the rodent lineage and 1,460 Kb to the bat lineage (Figure 4.4).

4.2.2.2 Charlie3 family

Charlie3 is the autonomous element that likely gave rise to the anthropoid-specific MITEs MER1A and MER1B. These TEs were previously classified as anthropoid-specific using the same methods used to classify Charlie12 (Pace and Feschotte 2007). Masking of all genomes with the human Charlie3 consensus revealed that a lineage-specific variant of Charlie3 was present not only in anthropoid primates, but also in bushbaby, tree shrew, cow and guinea pig. As with Charlie12, Charlie3 successfully invaded the anthropoid primate, bushbaby and guinea pig lineages, but also invaded the tree shrew and cow lineages. Unlike Charlie12, Charlie3 was unable to successfully infiltrate the murine rodent or bat lineages.

In both bushbaby and cow, a species-specific consensus sequence could be derived for Charlie3 (Charlie3_Og and Charlie3_Bt, respectively). Both consensuses had at least 92% nucleotide identity with the human Charlie3 over the entire length of the sequence. While we were able to find full-length or close to full-length Charlie3 elements in each species, we were unable to confidently derive a consensus sequence in guinea pig due to the high level of

fragmentation and sequence divergence between copies (25.4%). In tree shrew, we were unable to identify a sufficient number of full-length or near full-length Charlie3 elements to construct a consensus sequence. Therefore, as a proxy, the human Charlie3 consensus was used to mask the guinea pig and tree shrew genomes.

Charlie3 spawned a total of 13 distinct MITE families in 4 of the 5 lineages (MER1A and MER1B in anthropoid primates, MER1A_Bt_1 and MER1A_Bt_2 in cow, MER1A_Og_1 to MER1A_Og_5 in bushbaby and MER1A_Tb_1 to MER1A_Tb_4 in treeshrew). Two subfamilies, MER1A and MER1B have been previously described (see Repbase Update (Jurka et al. 2005)). As with Charlie3, no distinct MITE families could be confidently indentified in guinea pig due to the high level of sequence divergence. As such, the human MER1A and MER1B consensus sequences were used to mask the guinea pig genome. All MITE families attained a copy number of at least 790, with MER1B in anthropoid primates achieving a copy number of over 5,900 (Appendix D).

The graph of Charlie3 activity differs from that of Charlie12 in that it appears that all invasions occurred at approximately the same time period (between 60 and 80 mya), rather than occurring at different times (Figure 4.6). Surprisingly, the invasions into the three primate lineages (anthropoids, bushbaby and tree shrew) appear to be independent, occurring after the divergence of these species.

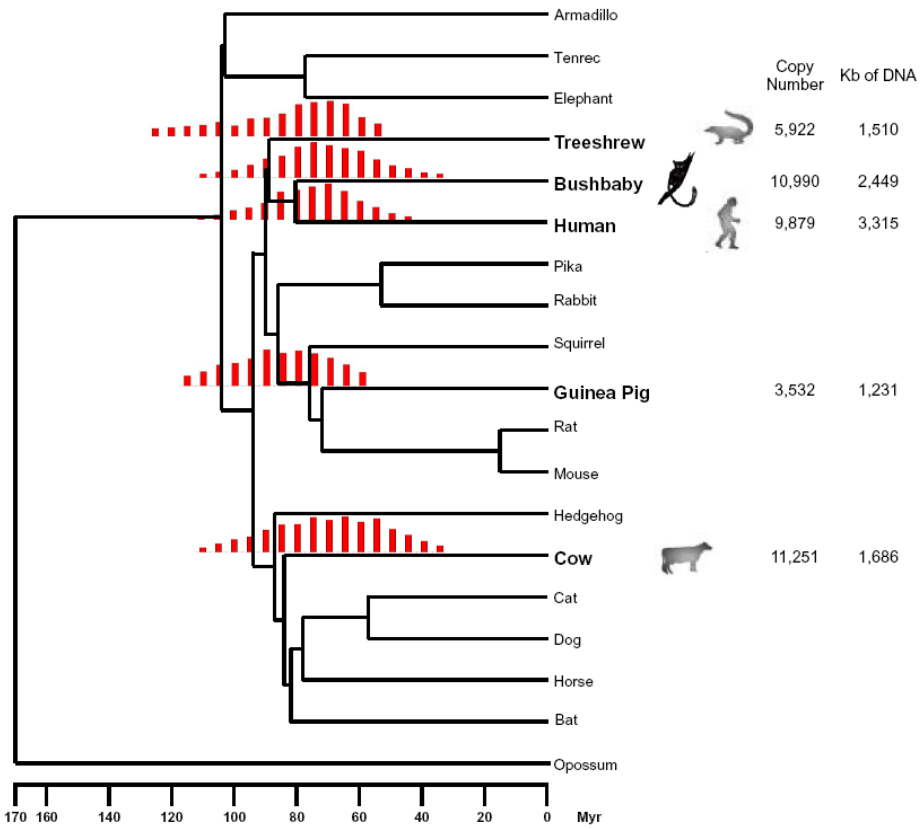


Figure 4.6 Species distribution and timing of amplification of Charlie3 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring Charlie3 tribe transposons are in bold. The timing of Charlie3 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all Charlie3 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

4.2.2.2.1 Charlie3 in primates

Using CSA, no copies of human Charlie3 tribe elements were identified at orthologous loci in bushbaby or tree shrew. Conversely, no copies of bushbaby-specific Charlie3 family elements could be found in human or tree shrew, nor could tree shrew-specific elements be found bushbaby or human. Instead, PIES could be found for each type of element between species (Figure 4.7a-c).

```
MER1A empty site in bushbaby
Human position:     chr2:234529045-234529731
Bushbaby position: scaffold_101604.1-169917:88236-88678
Human:              gtcttagagcaggggtgccgg...tgactgctgttttagagccaga
Bushbaby:           gtcctagag------------...------------------ctagg
```

(a)

```
MER1A_Tb_2 site in human
Treeshrew position:  scaffold_72242.1-4479:3116-3816
Human position:      chr20:22880669-22881000
Treeshrew:           AACATGGGTGGCTT-AAAGCAGG...GGGACTTCAGACTTAAAGGAGCCTTCA
Human:               GACATGAGTGGCTTCAAA------...------------------GATCCTTCA
```

(b)

```
MER1A_Og_1 site in human
Bushbaby position:   scaffold_1806.1-15139:9665-10419
Human position:      chr9:100009919-100010323
Bushbaby:            AGTGGGTTTCATTTAAAGCTATggtctc...ggcattgctgCTTTAAATCAAATGATTC
Human:               CATGGGTTTCACTGAAAG----------...------------------CGAATGATTC
```

(c)

Figure 4.7 Pre-insertion empty sites in (a) bushbaby for human MER1A, (b) human for tree shrew MER1A_Tb_2 and (c) human for bushbaby MER1A_Og_1 element. Target site duplications are in bold and underlined.

Previous NIA for the Charlie3 tribe in humans demonstrated that MER1A and MER1B had inserted into primate-specific LINE and SINE elements (Pace and Feschotte 2007). NIA in bushbaby ranked the MER1A_Og_x MITEs as the next oldest elements after MER30_Og_1, which was shown above to have had lineage-specific activity in bushbaby. In tree shrew, NIA provided relatively poor resolution for dating of TEs since only DNA TEs were annotated and few insertions of DNA TEs into other DNA TEs occurred. From the available data, MER1A_Tb_3 was ranked as the youngest element. However, this result should be viewed cautiously given the small number of TEs ranked. Therefore, in this case, the NIA data does not provide additional support for the lineage-specific activity of Charlie3 in treeshrew.

The average age of the Charlie3 tribes of elements in primates all post-date the divergence of the species from the common ancestor (Appendix D). In human, the MER1A and MER1B MITEs are 67 and 78 myr, respectively. The bushbaby MITEs (MER1A_Og_1 to

MER1A_Og_5) range from 67 to 78 myr and the treeshrew MITEs (MER1A_Tb_1 to MER1A_Tb_4) range from 73 to 85 myr. If the primate radiation occurred approximately 89 mya, then the age of these families corroborates the CSA and NIA results that the elements are indeed lineage-specific and that the invasions of the Charlie3 families were independent.

4.2.2.2.2 Charlie3 in guinea pig and cow

The infiltrations of the Charlie3 family into the guinea pig and cow lineages appear to be independent as well. No Charlie3 elements were found in other rodent or Laurasiatherian species and thus, CSA revealed no orthologous copies of any guinea pig Charlie3 tribe elements in any other rodent species, or any cow-specific elements in other Laurasiatherian species. Instead, PIES for cow-specific elements could be found in dog (Figure 4.8). As before, PIES could not be identified between guinea pig and other rodent species.

```
MER1A_Bt_1 empty site in dog
Cow position:  chr2:103843236-103844056
Dog position:  chr25:41753145-41753522
Cow:           TTTGTTATCTTTGTAGAGCAGGGGTCCC...GGCACCACTGCTCTAGAGCACATCTTAT
Dog:           TTTGTT-TCTCTGCTGAG----------...------------------CACATCTTAC
```

Figure 4.8 Pre-insertion empty site in dog for cow MER1A_Bt_1 element. Target site duplications are in bold and underlined.

The calculated ages of the Charlie3 tribe elements further suggest that they are indeed lineage-specific. In cow, the age of the Charlie3_Bt (60 myr), MER1A_Bt_1 (77 myr) and MER1A_Bt_2 (68 myr) elements all postdate the divergence of cow from the Laurasiatherian lineage (84 mya). In guinea pig, the families range in age from 77 myr (Charlie3) to 93 myr (MER1B). The dates of the MITE elements in guinea pig are significantly higher than the ages of the MITE families in the other species. If these elements are truly this old in guinea pig, then we would expect to find them in other rodents, but this is not the case. As before, this may be the result of not having an accurate molecular clock for this period of rodent evolution to use to date the elements. In addition, the elevated dates may be due to the lack of a confident consensus sequence for the elements in guinea pig and the necessity of masking the genome

with human consensus sequences. Regardless, the ages support the conclusion that the cow Charlie3 tribe elements are lineage-specific.

The final evidence for the lineage-specific activity of the cow Charlie12 tribe is provided by NIA. MER1A_Bt_1 (77 myr) was found nested within Tigger1 (64 myr), Tigger2f (67 myr) and MER1A_Bt_2 (68 myr). The oldest element that was nested within a MER1A_Bt_1 was MER1A_Bt_2 and vice versa. The youngest elements that MER1A_Bt_2 inserted in were L1MC1 (75 myr) and MER1A_Bt_1. Therefore, we can place the age range of these elements between approximately 64 and 77 mya. The output from TCF ranked MER1A_Bt_1 and MER1A_Bt_2 consecutively, with MER1A_Bt_2 first. Charlie3_Bt was ranked two elements older than MER1A_Bt_1. In addition, all three families were ranked younger than L1_Art (87 myr) and older than BovB (28 myr). Therefore, NIA does not contradict the conclusion that Charlie3_Bt activity is indeed cow-specific.

As with tree shrew, little data is provided for guinea pig from NIA. While Charlie3 is ranked as the youngest element, there are only a total of 14 elements that met TCF's connectedness criteria and were analyzed. Therefore, due to the dearth of usable data, no conclusions can be drawn from NIA in guinea pig.

The Charlie3 family of TEs was responsible for intense lineage-specific activity in the same major clades as Charlie12 (primates, Rodentia and Laurasiatheria), but in different species. This activity was more spectacular than that of Charlie12, generating almost twice as many lineage-specific MITE families (13 in Charlie3 versus 7 in Charlie12). Overall, these elements contributed 3.3 Mb of DNA to the anthropoid primate lineage, 2.4 Mb to the prosimian lineage, 1.5 Mb to the scandentia lineage, 1.2 Mb to the guinea pig lineage and 1.7 Mb to the cow lineage (Figure 4.6).

4.2.2.3 OposCharlie1 family

OposCharlie1 is an autonomous element that was annotated in the opossum genome, but was previously unannotated in any placental mammal genomes. Repbase Update

describes this element as opossum-specific. However, this element was also identified in tenrec, bushbaby and bat after each of the placental mammal genomes were masked with the opossum OposCharlie1 consensus sequence. The element was not found in any of the other mammalian species. OposCharlie1 is a relatively low copy number element, with only 148 copies in bat, 168 in bushbaby and 517 in tenrec (Appendix D). All copies are highly fragmented and no full-length copies could be recovered. However, there are over 16,000 copies in opossum, with 65 of these being full-length. The lack of full-length elements in the 3 placental mammal species prevented the confident reconstruction of the species-specific consensus sequences. Therefore, the consensus for the opossum OposCharlie1 element was used for masking all placental mammals.

In the bushbaby, tenrec and bat genomes, 7 different OposCharlie1 MITE families were propagated (OposCharlie1_NA_1 in tenrec, OposCharlie1_NA_2_Og in bushbaby, and OposCharlie1_NA_3_Ml to OposCharlie1_NA_6_Ml and nhAT2_Ml in bat). All 6 of these MITE families were previously undescribed, except nhAT2_Ml (Appendix C). Ray et al. described the non-autonomous element nhAT2_Ml which is highly similar to OposCharlie1, but whose 3' end differs (Ray et al. 2007). However, 162 of the 180 bp of nhAT2_Ml match the OposCharlie1 consensus sequence with 94% nucleotide identity. Therefore, this element is included as a MITE in the bat OposCharlie1 tribe.

The activity of the OposCharlie1 in the different species clearly shows a pattern reminiscent of SPIN elements (Figure 4.9). Like SPIN, there appears to be two different waves of OposCharlie1 infiltration. The first invasion occurred approximately 28-38 mya in the tenrec and bushbaby lineages and the second near 13 mya in bat. While recent invasions of DNA transposons into the bat genome have been previously described (Ray et al. 2008; Ray et al. 2007; Pritham and Feschotte 2007), recent invasions in other mammals have only been described for SPIN.

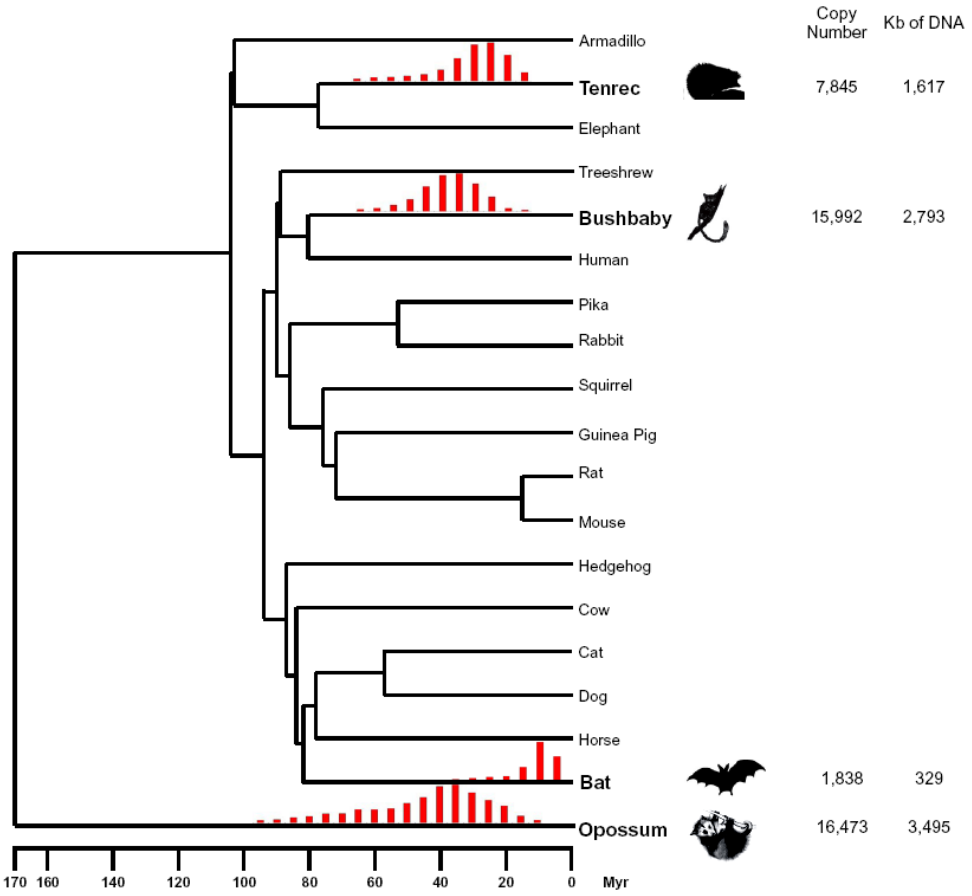


Figure 4.9 Species distribution and timing of amplification of OposCharlie1 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring OposCharlie1 tribe transposons are in bold. The timing of OposCharlie1 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all OposCharlie1 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

4.2.2.3.1 OposCharlie1 in bushbaby

SPIN likely infiltrated the bushbaby and tenrec genomes between 25 and 44 mya. The average divergence of bushbaby OposCharlie1 tribe elements range from 9.2 to 10.6%, within the range of bushbaby SPIN tribe elements (range = 7.5-13.0%), indicating that these two tribes were most likely transpositionally active during the same time period. This conclusion is further supported by the dating of these families. The average ages of OposCharlie1 and

OposCharlie1_NA_2_Og in bushbaby are 31 and 33 myr, respectively, with the SPIN families ranging from 25-44 myr.

CSA and NIA further support the age of the OposCharlie1 tribe elements in bushbaby. No orthologous copies of OposCharlie1 tribe elements could be identified in closely related species (Figure 4.10). NIA ranked OposCharlie1_NA_2_Og as the youngest family of elements in bushbaby, with the SPIN families as the next youngest. Since the rankings of element families by TCF are relative and these families are ranked in succession, we cannot absolutely conclude that the OposCharlie1 tribe is younger than SPIN in bushbaby. However, we can reasonably conclude from the data that their ages are approximately the same. OposCharlie1_NA_2_Og elements were found nested within SPIN_NA_1_Og elements, but the converse was not true. However, no DNA TEs were found nested within OposCharlie1_NA_2_Og elements, so this data cannot necessarily be interpreted to mean that there was no overlap in the activity of the OposCharlie1 and SPIN tribes. Based upon the ages, CSA and NIA, it is reasonable to conclude that the bushbaby OposCharlie1 tribe of elements are approximately 31-33 myr old and were active during the same time period as SPIN.

```
OposCharlie1_NA_2_Og site in bat
Bushbaby position:  scaffold_24286.1-1964:35-631
Bat position:       gi|105193115|gb|AAPE01402544.1:1035-1340
Bushbaby:           ATATACGCAACTAGACCAGGGGTCCT...AGGATGCCTGAACTAGACAATGATGCC
Bat:                ATATAATCAACTAGAT----------...----------------AAT---CCC
```

Figure 4.10 Pre-insertion empty site in bat for bushbaby OposCharlie1_Na_2_Og element. Target site duplications are in bold and underlined.

4.2.2.3.2 OposCharlie1 in tenrec

The average divergence of tenrec OposCharlie1 tribe elements ranges from 8.2 to 11.0%. Interestingly, the divergence of OposCharlie1_NA_1_Et from the consensus sequence in tenrec is only 8.2% as compared to an average divergence of 10.1 to 11.2% for tenrec SPIN tribe elements. Thus, the average age of OposCharlie1_NA_1_Et is only 28 myr, while the SPIN families range in age from 34-39 myr. These data suggest that OposCharlie1_NA_1_Et may be somewhat younger than SPIN in tenrec as in bushbaby.

NIA ranked OposCharlie1 and OposCharlie1_NA_1_Et as the first and third youngest elements in the tenrec genome, respectively, with SPIN_NA_1_Et as the fourth youngest. Six instances of OposCharlie1_NA_1_Et inserting into SPIN_NA_6_Et were found, and one instance of the converse was found. Since SPIN_NA_6_Et was found nested in an OposCharlie1_NA_1_Et element, we must conclude that the transpositional activity of both elements overlapped for a period of time. However, based on the calculated ages of the elements, OposCharlie1_NA_1_Et may have been active subsequent to the demise of SPIN.

4.2.2.3.3 OposCharlie1 in bat

In bat, the OposCharlie1 family appears to have infiltrated the lineage more recently than in tenrec or bushbaby. As with the OposCharlie1 tribes in bushbaby and tenrec, the average divergence and dates of the bat OposCharlie1 tribe elements are within those of the SPIN tribe. The OposCharlie1 tribe elements range from 3.5 to 6.3% divergence and 13 to 23 myr old, while those of SPIN range from 3.2 to 5.6% and 12 to 21 myr old.

The OposCharlie1 tribe elements were classified as approximately the same age as SPIN elements by NIA. nhAT2 was ranked as the sixth youngest element, just before SPIN_NA_9_MI and 5 families before the other SPIN MITEs. These data suggest that the bat OposCharlie1 tribe elements were contemporary with SPIN and were active at the same time, thereby providing additional support for the calculated ages. CSA revealed that no orthologous copies of OposCharlie1 tribe elements in bat could be found in bushbaby or tenrec.

As with the other horizontally transferred hAT elements, each invasion appears to be independent. Using CSA, no orthologous copies of OposCharlie1 tribe elements could be located between bushbaby, tenrec or bat, but PIES could be identified (Figure 4.10). PIES could be found in closely related species as well (Figure 4.11). Given that the ages of all OposCharlie1 tribe elements in bat, bushbaby and tenrec postdate the divergence of the species from their common ancestor and that each family of OposCharlie1 tribe elements were

found inserted into known lineage-specific elements, we conclude that the infiltrations were indeed independent.

```
OposCharlie1_NA_3_Ml site in dog
Bat position:  gi|05929654gbAAPE01058739.1:1979-3098
Dog position:  chr10:15731325-15731664
Bat:           ATGATATATTATCTACTCCAGGGGTCC...AGGACCCCTGATCTACTCTAGGAGCACG
Dog:           ATGATTTATTCCCTAACC---------...------------------CAGCAGCATG
```

Figure 4.11 Pre-insertion empty site in dog for bat OposCharlie1_Na_3_Ml element. Target site duplications are in bold and underlined.

4.2.2.3.4 Evolutionary dynamics of OposCharlie1

The evolutionary dynamics of OposCharlie1 in the different genomes are striking. In bat, 5 distinct MITE families were derived from OposCharlie1. From these families, 7,741 copies were created. In comparison, only 1 MITE family was spawned in bushbaby and a separate single MITE family generated in tenrec. In bushbaby, the OposCharlie1_NA_2_Og MITE family attained a copy number of over 15,800, with almost 12,800 of these elements being full-length. Surprisingly, only 20 more fragments of the OposCharlie1 autonomous element are found in bushbaby (n=168) than in bat (n=148), yet over twice as many MITE copies were generated in the bushbaby lineage. Over 7,300 copies of the tenrec-specific MITE OposCharlie1_NA_1_Et could be identified, almost 5,000 of which are full-length. While there are over 3 times as many OposCharlie1 fragments in tenrec (n=517) as in either bushbaby or bat, there are less MITE copies than in any of the genomes.

The activity of OposCharlie1, in general, is similar to that of SPIN. Both infiltrated the bushbaby, bat, tenrec and opossum lineages independently and produced two apparently distinct bursts of activity. This seemingly concomitant activity raises the question about whether the vector that transferred SPIN into these species was also responsible for the introduction of OposCharlie1. Unfortunately, since the vector is currently unknown, this remains an untestable hypothesis.

4.2.2.4 RCHARR1 family

RCHARR1 is an “orphan element” that was previously classified as murine rodent specific (Waterston et al. 2002). After the placental mammal genomes were masked with the RCHARR1 consensus sequence from mouse, it was discovered that a variant of RCHARR1 was also found in tenrec. This variant, named RCHARR1_Et, was distinct from RCHARR1 in mouse in length (229 bp and 974 bp, respectively) (Appendix C). Neither element could be identified in any other mammalian species.

4.2.2.4.1 RCHARR1 in tenrec

The calculated age of RCHARR1_Et based on average divergence was 84 myr. This age slightly predates the divergence of tenrec from elephant (77 mya), suggesting that we might find orthologous copies of RCHARR1_Et in elephant. However, since the average age is close to the divergence date, and the divergence date is an estimated date, there was also an equal probability of not finding orthologous copies. Using CSA, we were unable to find any orthologous RCHARR1_Et copies between tenrec and elephant, but rather were able to identify PIES in elephant (Figure 4.12). This result validates the RepeatMasker output that found no copies of RCHARR1 in elephant. The relative age of RCHARR1_Et could not be determined from NIA because RCHARR1_Et did not meet the connectedness criteria in TCF. As such, we can conclude that RCHARR1_Et is indeed specific to the tenrec lineage.

```
RCHARR1_Et site in elephant
Tenrec position:     scaffold_318971:35603-36237
Elephant position:   scaffold_14432:34235-34599
Tenrec:              AAT-GT-TAC ATAAGACT CAGGACTTCT...AGAAGCTTAA AGCACTGC TGTGACTTTC
Elephant:            AATAGTATATA----------...----------  ACTGATGC TGTGTCCTTC
```

Figure 4.12 Pre-insertion empty site in elephant for tenrec RCHARR1_Et element. Target site duplications are in bold and underlined.

4.2.2.4.2 RCHARR1 in rodents

RCHARR1 in mouse was calculated to be 84 myr old. This age predates the divergence of both squirrel and guinea pig from the rodent lineage. However, RepeatMasker did not identify any copies of RCHARR1 in either squirrel or guinea pig. We were unable to

identify any PIES in guinea pig or squirrel due to the high level of sequence divergence between these species. Therefore, we can assume that RCHARR1 infiltrated the rodent genome after the split of guinea pig from the murine rodent lineage. This apparent conflict between the calculated age and CSA may again be the result of an overestimation of the age due to an incorrect neutral substitution rate for DNA TEs during this period of rodent evolution.

NIA ranked RCHARR1 as slightly older in age than MTEa (85 myr) and MTEb (86 myr) and slightly younger than MER2 (98 myr) and L1MA5 (101 myr). We were able to find RCHARR1 nested within MER2. Interestingly, we were able to find insertions of apparently much older TEs, such as the B4 SINE element (103 myr), inserted into RCHARR1. This may suggest that RCHARR1 was present in the rodent lineage prior to 100 mya but was lost in the guinea pig and squirrel lineages. The more likely scenario is that B4 is actually younger than 103 myr but the age is currently being overestimated.

Together, these data suggest that RCHARR1 and RCHARR1_Et were horizontally transferred independently into the murine rodent and tenrec lineages (Figure 4.13). Relative to other horizontally transferred TEs in the tenrec genome such as OposCharlie1_NA_1_Et and the SPIN MITEs, RCHARR1_Et was relatively unsuccessful in its colonization, generating only 1,223 copies. In contrast, over 6,400 copies of RCHARR1 are still recognizable in mouse and over 6,200 in rat. While RCHARR1 and RCHARR1_Et were responsible for lineage-specific activity in both murine rodents and tenrec, the genomic impact in the species was significantly different. In mouse, RCHARR1 contributed 1.9 Mb of DNA to the genome. In contrast, RCHARR1_Et contributed only one-tenth of the amount of DNA (194 Kb) to the tenrec genome (Figure 4.13).

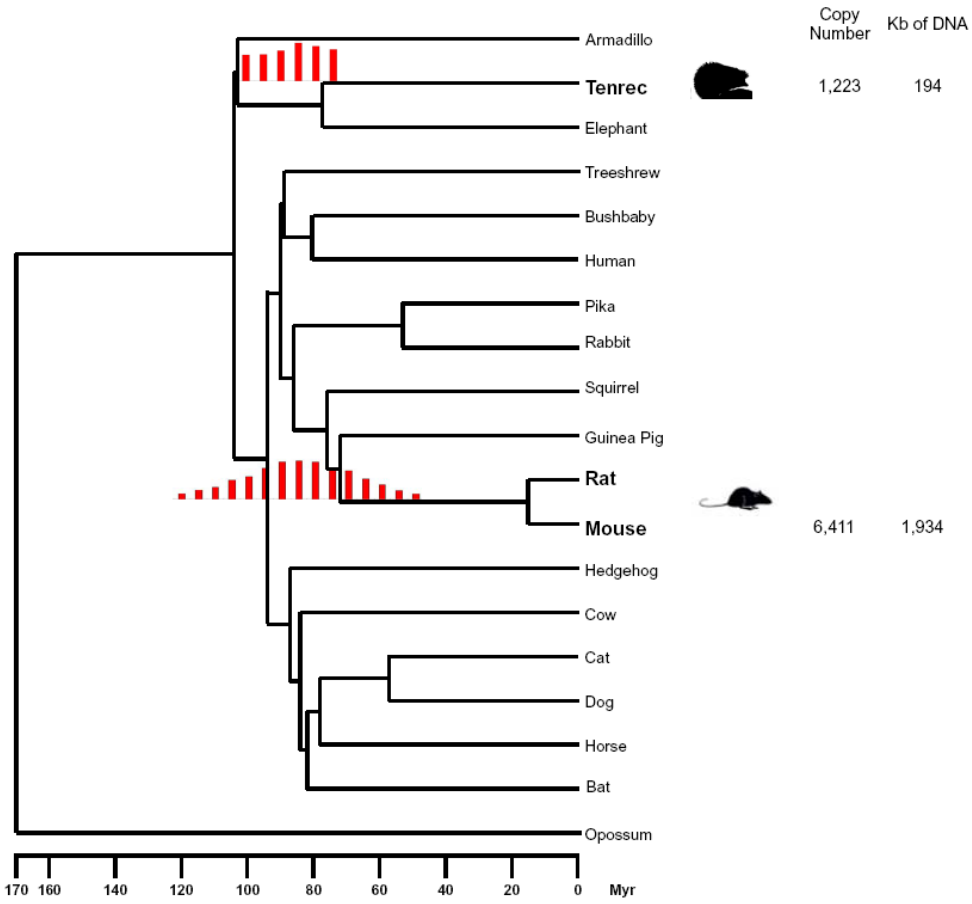


Figure 4.13 Species distribution and timing of amplification of RCHARR tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring RCHARR tribe transposons are in bold. The timing of RCHARR tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all RCHARR tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

4.2.3 Horizontal transfer of Tc1/mariner elements

A total of 7 different families of Tc1/mariner superfamily transposons have been independently horizontally transferred into mammalian species (Dumbo, HSMAR1, HSMAR2, LAMAR2, Oamar1, OGMAR1 and OGMAR2). All are autonomous elements that spawned distinct MITE families in each genome (Appendix C).

4.2.3.1 Dumbo family

Dumbo is a previously unidentified mariner element found only in Afrotherian species (tenrec and elephant). The element is 1,608 bp long and encodes a putative 292 amino acid transposase that has a conserved CENP_B N-terminal DNA binding domain and CENP_B DNA binding domain. The closest related Tc1/mariner element is Tigger3(Golem) which has ~50% identity at the amino acid level. Dumbo and LAMAR2 are the only currently known Afrotherian-specific DNA TEs. Two distinct MITE families originated from Dumbo: Dumbo_NA_1 and Dumbo_NA_2.

4.2.3.1.1 Dumbo in tenrec

CSA reveals that Dumbo invaded the Afrotherian ancestor prior to the divergence of elephant and tenrec as orthologous copies of all Dumbo tribe families could be identified in both species. RepeatMasker did not identify any copies of Dumbo in any other mammalian genomes, and this result is supported by PIES which were identified in related species (Figure 4.14). The calculated age of Dumbo in tenrec (91 myr), which is prior to the divergence of elephant and tenrec (77 mya), further corroborates the CSA data. Data from NIA suggests that Dumbo in tenrec is approximately as old as LAMAR2_Et (90 myr) (see below), another Afrotherian-specific DNA TE, and younger than MER20 (104 myr), a eutherian-wide DNA TE. This data, taken with the CSA and age data, clearly demonstrate that Dumbo was present in the common ancestor of elephant and tenrec and is clearly absent from all other mammals.

```
Dumbo_NA_1 site in human
Elephant position:   scaffold_22448:1451-2167
Human position:      chr11:97350668-97351050
Elephant:            TAAACAACAGTACAGGTAATCC...GATTACC-TATTTTCAATCCT
Human:               TAAACAAAAGTA----------...----------TTTCAGTTCC
```

Figure 4.14 Pre-insertion empty site in human for elephant Dumbo_NA_1 element. Target site duplications are in bold and underlined.

4.2.3.1.2 Dumbo in elephant

The graph of Dumbo tribe activity reveals an interesting pattern (Figure 4.15). According to the ages of the Dumbo families (range = 61 to 80 myr), it appears that the transpositional activity of Dumbo occurred not only in the ancestor of tenrec and elephant, but that significant lineage-specific activity occurred in elephant. However, no evidence could be found to support this conclusion, as no PIES for any Dumbo tribe elements could be identified in either species. Instead, dozens of orthologous copies could be found. Based upon this data, we must conclude that if any lineage-specific Dumbo activity occurred in either tenrec or elephant, it was minimal. The discrepancy in the ages of Dumbo tribe elements in tenrec and elephant is most likely due to an incorrect calibration of the molecular clock during this period. As the age of Dumbo in tenrec is corroborated by CSA and NIA, but not in elephant, then we must assume that the age of Dumbo in elephant is incorrect. This is most likely due to the neutral substitution rate in elephant since the time period when Dumbo was active being slower than the average rate we derived, implying that the neutral rate has increased in elephants at some point since that time.

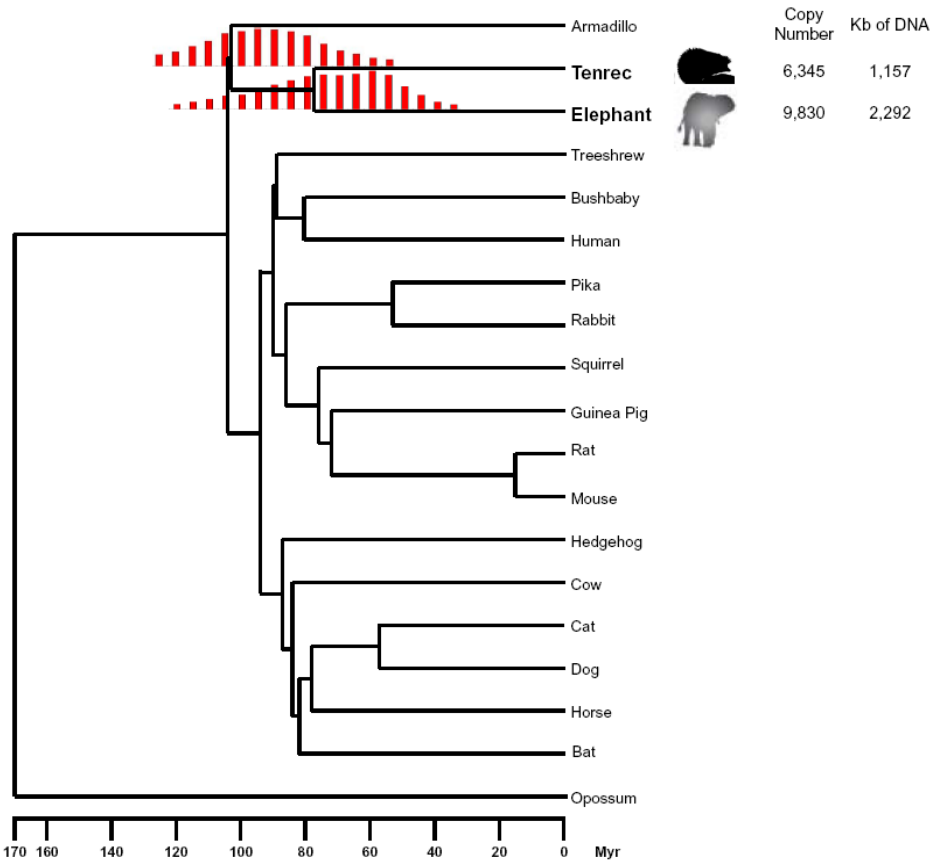


Figure 4.15 Species distribution and timing of amplification of DUMBO tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring DUMBO tribe transposons are in bold. The timing of DUMBO tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all DUMBO tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

4.2.3.2 HSMAR1 family

HSMAR1 is an autonomous Tc1/mariner element that was previously classified as anthropoid-specific (Pace and Feschotte 2007; Robertson and Zumpano 1997). This element has played an important role in anthropoid primate evolution, spawning the MADE1 MITE which amplified to approximately 8,000 copies. In addition, the SETMAR gene was formed by the fusion of an HSMAR1 element with a SET domain. This protein has been shown to bind a specific motif of MADE1 (Cordaux et al. 2006b).

Previous studies found that a murine rodent-specific element termed MMAR1 is 95% identical to HSMAR1 at the nucleotide level and that the presence of these elements in both genomes is most likely the result of independent horizontal transfers (Waterston et al. 2002). In order to investigate this claim, we masked the mammalian genomes and were able to detect HSMAR1 not only in anthropoid primates and mouse, but also in bushbaby, treeshrew, cow and armadillo. Lineage-specific consensus sequences were derived for each species (HSMAR1_Og, HSMAR1_Tb, HSMAR1_Bt and HSMAR1_Dn, respectively) and these were used to subsequently remask the genomes to obtain accurate divergence estimates. All consensus sequences had a minimum of 95% nucleotide identity with the human HSMAR1 consensus. Surprisingly, anthropoid primates were the only lineage in which HSMAR1 generated a MITE family.

4.2.3.2.1 HSMAR1 in primates

We began by trying to determine if the invasions of HSMAR1 into the 3 primate species were independent or if HSMAR1 was present in the common ancestor. The HSMAR1 family that is found in human has previously been demonstrated to be anthropoid-specific using CSA, NIA and dating of elements based on sequence divergence (Pace and Feschotte 2007). Therefore, we knew that HSMAR1 had independently invaded the ancestor of anthropoid primates and that all insertions were lineage-specific. This led us to believe that the same may be true for the bushbaby and tree shrew lineages.

To further investigate this possibility, CSA was performed to search for the presence of orthologous copies of HSMAR1 between the primate lineages, but no orthologs were found in any of the primate species. However, PIES could be identified (Figure 4.16a-b). We further checked for orthologous HSMAR1 elements between the primate species and all other non-primate mammalian species where HSMAR1 was identified, but no orthologs could be found either and PIES could be identified (Figure 4.16c). From these data, we can conclude that no copies of HSMAR1 in any of the three primate species, or any of the non-primate species, are

orthologous and therefore must be the result of independent infiltrations of HSMAR1 followed by lineage-specific activity.

```
HSMAR1 site in bushbaby
Human position:      chr6:128320441-128322083
Bushbaby position:   scaffold_85512.1-352477:57159 -57399
Human:               TGCATACTTCTGttaggttggt...actaacctaaTACCCCAGTAGT
Bushbaby:            TGTATGCTACTA----------...------------CTCCAACAGT
```

(a)

```
HSMAR1_Tb site in human
Tree shrew position: scaffold_11447.1-280991:40133-41793
Human position:      chr3:118647561-118647950
Tree shrew:          AATCCAAAAATATTAGGTTGAT...CCTAAATTAAGAGTAA-T-TA-T
Human:               AATTCAAAAATA----------...------------GTAACTGTACT
```

(b)

```
HSMAR1 site in cow
Human position:      chr6:11442590-11444266
Cow position:        chr23:42589737-42589984
Human:               T--CTGAAAATAttaggttgga...accaatctaaTATATTTCAAAG
Cow:                 TTTCAGAAAGGT----------...-------------CTTTCTGTA
```

(c)

Figure 4.16 Pre-insertion empty sites in (a) bushbaby for human HSMAR1, (b) human for tree shrew HSMAR1_Tb element and (c) cow for human HSMAR1 element. Target site duplications are in bold and underlined.

Further analysis of the activity of HSMAR1 in primates reveals that each infiltration appears to have occurred at approximately the same time, rather than in different time periods (range = 46 myr in armadillo to 63 myr in tree shrew) (Figure 4.17). In treeshrew, the average divergence of the HSMAR1_Tb family elements is 18.6%, resulting in a calculated age of 63 myr. This is the lowest average divergence of all DNA element families in tree shrew, including MER1A_Tb_1 (73 myr), the Charlie3 MITE that was determined to be tree shrew-specific above. As stated previously, NIA data for tree shrew is unreliable due to the dearth of DNA element insertions into other DNA elements.

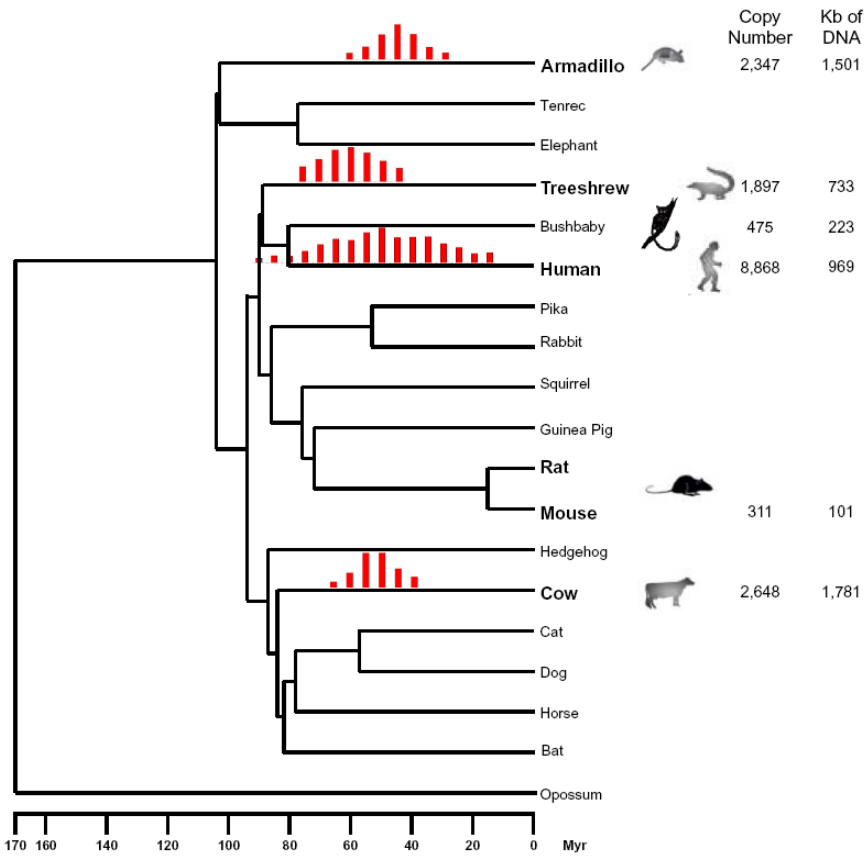


Figure 4.17 Species distribution and timing of amplification of HSMAR1 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring HSMAR1 tribe transposons are in bold. The timing of HSMAR1 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all HSMAR1 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

The calculated age of HSMAR1 in bushbaby is 59 myr, with an average divergence of 17.5%. While this is not the youngest family of elements in bushbaby, it is younger than the Charlie3 and Charlie12 tribe elements, both of which were demonstrated to be bushbaby-specific. In addition, it is approximately the same age as the bushbaby-specific OGMAR1 and OGMAR2 families (both 60 myr) (see below). Because of the low copy number of HSMAR1_Og elements in bushbaby (n=475), it did not meet the connectedness criteria for TCF and was not

ranked in age based on nestings. The average age of HSMAR1 elements in human, as determined previously, is 56 myr.

In all primate species, the average age of the HSMAR1 elements postdate the primate radiation (89 mya). Therefore, we can confidently claim that HSMAR1 independently invaded each of the tree shrew, bushbaby and anthropoid primate lineages approximately 55-65 mya, producing differential lineage-specific activity in each species.

4.2.3.2.2 HSMAR1 in rodents

To determine the time of the MMAR1 invasion into the rodent lineage, we again employed CSA. No copies of MMAR1 were masked in guinea pig or squirrel, two additional rodent species, so we began with the assumption that MMAR1 had invaded the rodent lineage subsequent to the divergence of murine rodents from squirrel or guinea pig (72 mya) We were able to identify multiple orthologous copies of MMAR1 between rat and mouse, demonstrating that the element was active prior to the divergence of these species ~23 mya. We were unable to identify any MMAR1 PIES in guinea pig due to the high substitution rates in both guinea pig and mouse.

Using NIA, we discovered that MMAR1 nested within 1 copy of MTEa, an LTR element that is calculated to be 88 myr, while suffering multiple insertions of RSINE1 (101 myr). However, MMAR1 was not ranked by TCF since it did not meet the connectedness criteria. From these data, we can conclude that MMAR1 is no older than MTEa and no younger than RSINE1. As with other DNA TEs that have been horizontally transferred into the rodent lineage, the average age of MMAR1 (84 myr) significantly predates the divergence of murine rodents from guinea pig or squirrel. Again, because of the uncertainty of the neutral substitution rate in rodents during this time period, this date may well be an overestimate of the actual date. Based upon the absence of MMAR1 in guinea pig and squirrel, it is reasonable to assume that MMAR1 invaded the rodent lineage after the divergence of murine rodents from guinea pig, or within the last 72 myr. However, without more accurate dating of this element or further NIA data, we

cannot exclude the possibility that MMAR1 was present in the common ancestor of these species, but was subsequently lost in the guinea pig and/or squirrel lineage.

4.2.3.2.3 HSMAR1 in cow

In the Laurasiatheria clade, HSMAR1 was only able to invade the cow lineage, as no copies were identified by RepeatMasker in any other Laurasiatherian genomes. In cow, the average divergence of HSMAR1_Bt copies from their consensus sequence was only 14.0%. This divergence translates to a date of approximately 50 myr. TCF ranked HSMAR1_Bt as approximately the same age as OAMAR1, a cow-specific DNA TE that is 53 myr old, and younger than Tigger1a_Car (75 myr), L1MA6 (99 myr) and MER2 (83 myr). HSMAR1_Bt was found nested within ART2A, a cow-specific SINE element whose age is calculated to be 51 myr, while the oldest TE nested within HSMAR1_Bt was also ART2A. These data suggest that HSMAR1_Bt is indeed approximately 50 myr, an age that postdates the divergence of cow from the Laurasiatherian ancestor (84 mya), thereby making the activity of HSMAR1_Bt cow-specific.

CSA further corroborates the findings that HSMAR1_Bt is cow-specific, while also providing evidence that the infiltration was independent of the invasions of HSMAR1 into other species. PIES could be identified in closely related species (Figure 4.18), demonstrating that the insertions occurred only in cow. Additionally, no non-nested copies of any HSMAR1_Bt elements could be found at orthologous loci in any of the other species where HSMAR1 is present. Taken together with the age of HSMAR1_Bt and the NIA data, CSA allows us to confidently conclude that HSMAR1_Bt independently infiltrated the cow lineage, resulting in lineage-specific activity.

```
HSMAR1_Bt site in dog
Cow position:      chr1:58659035-58660696
Dog position:      chr33:26371901-26372290
Cow:               TATTGGTTAGTATTAGGTTGGT...ACCAACCTAATAGGTGTTCTCC
Dog:               TATAGGTTAGTA----------...------------GATGTTCTCA
```

Figure 4.18 Pre-insertion empty sites in dog for cow HSMAR1_Bt element. Target site duplications are in bold and underlined.

4.2.3.2.4 HSMAR1 in armadillo

HSMAR1 is the only one of the horizontally transferred elements we identified that was able to infiltrate the armadillo lineage. The calculated age of HSMAR1_Dn, the armadillo-specific variant of HSMAR1, indicates that it is approximately 46 myr old. Given that the average divergence of HSMAR1_Dn copies to their consensus sequence is only 12.2%, we believe this is a realistic estimate of the age. This age ranks HSMAR1_Dn as the youngest currently known DNA element in armadillo, approximately 16 myr younger than Tigger1 (64 myr).

CSA was used to confirm that HSMAR1_Dn was indeed specific to the armadillo lineage. No orthologous copies of non-nested HSMAR1_Dn elements could be identified in related species, and no non-nested copies in armadillo were found any other species where HSMAR1 was found. Instead, PIES could be identified (Figure 4.19). Based upon the age of HSMAR1_Dn in armadillo and the lack of any identifiable orthologs in other species, we conclude that the infiltration of HSMAR1_Dn into armadillo was independent of the invasions in other species. This invasion subsequently produced lineage-specific activity of the HSMAR1_Dn family of elements.

```
HSMAR1_Dn site in human
Armadillo position:     scaffold_65916:3519-5169
Human position:         chr1:48298987-48299362
Armadillo:              GAACAGATATTATTAGATTGGT...ACCAACCTAATAATTATGG-CAT
Human:                  caacatatgtaa----------...----------gttttGGGCAA
```

Figure 4.19 Pre-insertion empty sites in human for armadillo HSMAR1_Dn element. Target site duplications are in bold and underlined.

4.2.3.2.5 Evolutionary dynamics of HSMAR1 in mammalian species

Within the primate clade, HSMAR1 and its species-specific variants differed significantly in their level of activity. In human, 865 copies of HSMAR1 can be identified plus over 8,000 copies of MADE1. However, in bushbaby, only 475 copies of HSMAR1_Og are present and no corresponding MITE families were produced. In tree shrew, almost 1,900 copies of HSMAR1_Tb are masked, with no related MITE families. The reason for this

differential success in primates is unclear. The amplification of HSMAR1 in cow and armadillo (n= 2,648 and 2,347 copies, respectively) is slightly more than that of tree shrew. In mouse, MMAR1 only achieved a copy number of 331. In total, the HSMAR1 family added a total of 969 kb of DNA to the anthropoid primate, 223 kb to the prosimian (bushbaby), 733 kb to the scandentia (tree shrew), 1,781 kb to the cow and 1,501 kb to the armadillo lineages (Figure 4.17).

The discovery that HSMAR1 MITEs are only found in anthropoid primates is interesting. The reason for this may be connected with the domestication of the HSMAR1 transposase to form the SETMAR gene. If the dispersal of MADE1 throughout the anthropoid genomes was part of the creation of a gene regulatory network, as has been proposed (Feschotte 2008), then the removal of new MADE1 elements may have been slowed due to selection to maintain and build the regulatory network. Since HSMAR1 has not been domesticated in any other lineages, there would most likely be little, if any, selection to maintain HSMAR1 variants. Indeed, the new insertions may have been deleterious and would have been selected against.

4.2.3.3 HSMAR2 family

HSMAR2 is an autonomous Tc1/mariner element that was previously determined to be primate-specific (Pace and Feschotte 2007; Robertson and Martos 1997). After masking the mammalian genomes with the human HSMAR2 consensus sequence, we discovered that species-specific variants are also found in tenrec, treeshrew, pika and rabbit, but not in any other species. We were able to reconstruct species-specific consensus sequences for all species except pika and remasked the genomes using these sequences to obtain accurate divergence data (tenrec = HSMAR2_Et, treeshrew = HSMAR2_Tb, rabbit = HSMAR2_Oc). For pika, the human HSMAR2 sequence was used. In all cases, the species-specific HSMAR2 consensus sequences had at least 96% nucleotide identity with the human HSMAR2 consensus over the entire length of the element.

4.2.3.3.1 HSMAR2 in primates

We next sought to discover when the invasions of HSMAR2 occurred and if the invasions were independent. Using CSA, we verified that multiple copies of HSMAR2 that are present in human are also present at the orthologous locus in bushbaby. However, no orthologous copies could be indentified in tree shrew (Figure 4.20a). Additionally, no tree shrew HSMAR2 copies could be found at the orthologous locus in human or bushbaby, but PIES could be identified (Figure 4.20b). No human, bushbaby or tree shrew copies could be found at orthologous loci in any other species in which HSMAR2 is found as well (Figure 4.20c). From this data, we conclude that HSMAR2 invaded the primate genome two times independently, once in the tree shrew lineage and once in the time period between the divergence of tree shrews from the primate lineage and the divergence of prosimians from anthropoid primates.

```
HSMAR2 site in treeshrew
Human position:     chr2:20023950-20025649
Treeshrew position: scaffold_123380.1-58649:11010 -11387
Human:              TCATCATTTATAcaaggggtct...agtcccctcATATTGCTACAGT
Treeshrew:          CCATCATTTATA----------...------------CTTCTAAAAT
```

(a)

```
HSMAR2_Tb site in human
Treeshrew position:  scaffold_122477.1-143590:64479-66112
Human position:      chr7:135686688-135687058
Treeshrew:           GACAATTATGTATGAGGGGAGG...AGTCTCCTTAATTGTCTTCTCT
Human:               gcctatgacgca----------...------------tgtctgttcc
```

(b)

```
HSMAR2 site in cow
Human position: chr2:103466965-103468487
Cow position:   chr11:8323859-8324252
Human:          AGAACAACCATATgagggatct...agttcccttgTAGAAGATTTTT
Cow:            TGGACAAGCACG----------...------------CGAGGTTTTT
```

(c)

Figure 4.20 Pre-insertion empty sites in (a) treeshrew for human HSMAR2, (b) human for tree shrew HSMAR2_Tb element and (c) cow for human HSMAR2 element. Target site duplications are in bold and underlined.

Dating of HSMAR2 in human, bushbaby and tree shrew further supports the claim of independent infiltrations. In tree shrew, HSMAR2_Tb and HSMAR2_NA_1_Tb, an HSMAR2_Tb MITE, have calculated ages of 68 and 65 myr. In human and bushbaby, HSMAR2 is 61 and 65 myr, respectively. These dates all postdate the divergence of these species (89 myr).

4.2.3.3.2 HSMAR2 in tenrec

According to the calculated age, HSMAR2 appears to have invaded the tenrec lineage approximately 80 mya. The lineage-specific variant, HSMAR2_Et, spawned a single MITE family, HSMAR2_NA_4_Et, that reached a copy of over 42,000. If HSMAR2_Et and its corresponding MITE are truly 80 myr old, then, given that the divergence of these species was 77 mya, we expect to find copies of HSMAR2_Et in elephant.

To investigate this possibility, we used CSA to search for orthologous copies of HSMAR2_Et in elephant, even though RepeatMasker had not identified any copies of HSMAR2 in elephant when the genome was masked with the human HSMAR2 consensus sequence. This result was validated when we were unable to find any orthologous copies of HSMAR2_Et or HSMAR2_NA_4_Et in elephant, but instead were able to locate PIES (Figure 4.21). We were also unable to identify any orthologous copies of HSMAR2_Et in any other species where HSMAR2 is found. Therefore, we can confidently conclude that HSMAR2 is indeed absent from elephant and that the infiltration of HSMAR2 into tenrec was an independent event that occurred after the divergence of tenrec from elephant.

```
HSMAR2_NA_4_Et site in elephant
Tenrec position:    scaffold_300632:88320-88799
Elephant position: scaffold_28970:10651 -10992
Tenrec:             GTGCAGAAAATCTGAGGGGCAT...AGCCCCTTCTTTCATTTAGCTT
Elephant:           GTGCAGAAAATA----------...-----------TATTTAGTTT
```

Figure 4.21 Pre-insertion empty sites in elephant for tenrec HSMAR2_NA_4_Et element. Target site duplications are in bold and underlined.

4.2.3.3.3 HSMAR2 in lagomorphs

HSMAR2 is found in both of the lagomorph species surveyed (rabbit and pika). In both species, one MITE was spawned, HSMAR2_NA_2_Op and HSMAR2_NA_3_Oc. The average age of HSMAR2 in each species was approximately the same, with pika ranging from 67 to 81 myr for HSMAR2_NA_2_Op and HSMAR2, respectively. The ages of HSMAR2_Oc and HSMAR2_NA_3_Oc were 72 and 63 myr, respectively. These ages predate the divergence of pika from rabbit, suggesting that HSMAR2 invaded the common ancestor.

To determine if HSMAR2 did indeed infiltrate the common ancestor, we used CSA to search for orthologous copies in both species. We were able to identify dozens of copies of HSMAR2_Oc, HSMAR2_NA_2_Op and HSMAR2_NA_3_Oc that were shared between the species. We were unable to identify any copies of lagomorph-specific HSMAR2 in the other species where HSMAR2 is present. Therefore, we conclude that HSMAR2 independently invaded the common ancestor of rabbit and pika approximately 70 mya (Figure 4.22).

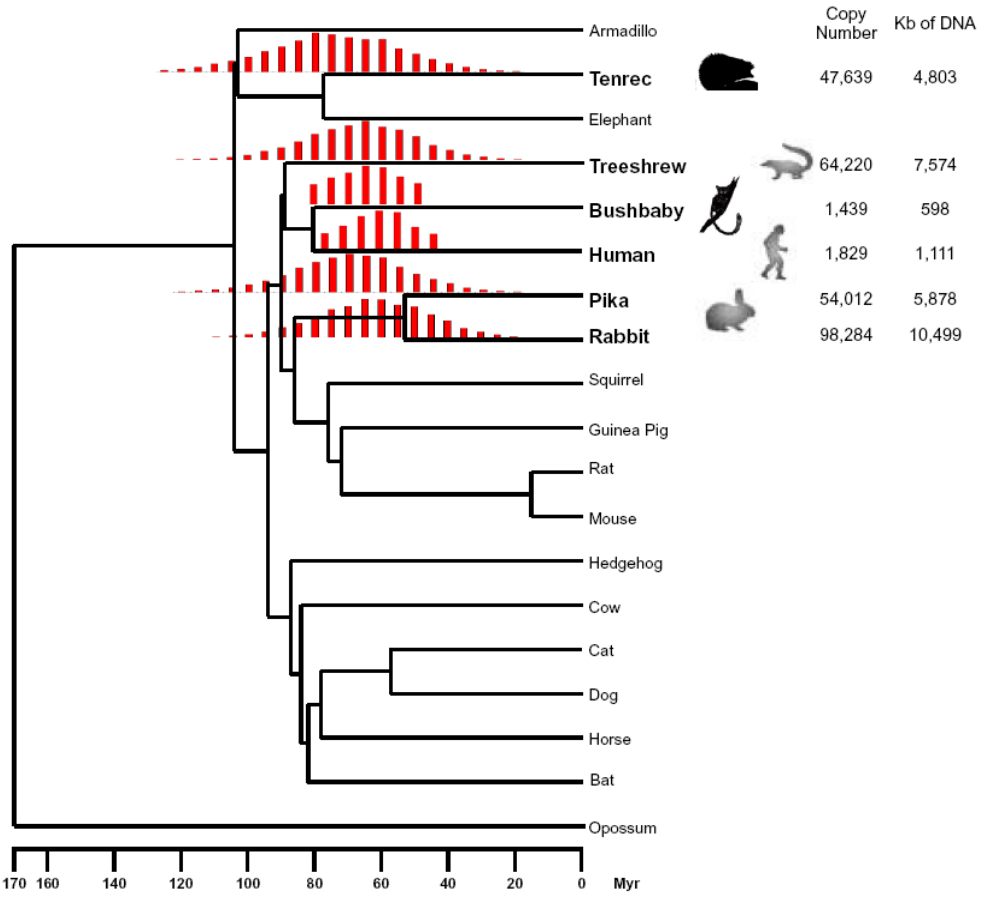


Figure 4.22 Species distribution and timing of amplification of HSMAR2 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring HSMAR2 tribe transposons are in bold. The timing of HSMAR2 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all HSMAR2 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

4.2.3.3.4 Evolutionary dynamics of HSMAR2

HSMAR2 appears to have invaded all lineages at approximately the same time. This suggests that there may have only been one vector that carried and transferred HSMAR2 to all of the species. This possibility is intriguing and suggests that there may have been geographical overlap between the species in Africa.

In all species except human and bushbaby, HSMAR2 spawned one lineage-specific MITE. In tree shrew, HSMAR2_NA_1_Tb reached a copy number of over 54,000, making it the highest copy number element in tree shrew, with more than 5 times as many copies as MER5A, the second highest copy number element. HSMAR2_Tb itself attained a copy number of over 9,800. In contrast, the human and bushbaby HSMAR2 only reached ~1,700 and 1,400 copies, respectively.

In lagomorphs, the pika HSMAR2_NA_2_Op and rabbit HSMAR2_NA_3_Oc also attained remarkably high copy numbers (~46,000 and ~86,000, respectively) with HSMAR2_Op and HSMAR2_Oc reaching copy numbers of greater than 7,900 in both lineages. In both of these species, the HSMAR2 MITE is the highest copy number element. The same scenario is true in tenrec where HSMAR2_NA_4_Et multiplied to over 42,000 copies and HSMAR2_Et to 5,500 copies. The only element to reach higher copy number in tenrec is SPIN_NA_1_Et (n=52,541).

While copies of the HSMAR2 MITEs in lagomorphs inserted before the divergence of these species and thus can be found at orthologous loci, the copy number in rabbit is almost double that of pika. This may be explained in two ways. First, there may have been significant lineage-specific activity of the HSMAR2_NA_3_Oc MITE in rabbit after the divergence of these species, but we were unable to verify this as no clear empty sites in pika could be identified. Second, a survey of the copy numbers of all shared DNA TE families in rabbit and pika revealed that, on average, rabbit has 1.8 times more copies than pika. Therefore, the pika genome is most likely more adept at the removal of transposable elements than rabbit, thus creating the large difference in copy number.

One intriguing finding about HSMAR2 is that no MITE family was spawned in human, and that HSMAR2 achieved a relatively meager copy number in anthropoid primates as compared to the other species. While primates were by no means immune from the horizontal transfer of DNA transposons followed by massive bursts of activity, such as in the case of

HSMAR1 with MADE1, HSMAR2 was unable to achieve high copy number. One possible explanation for the low copy number of HSMAR2 is that some defense mechanism may have evolved to suppress its activity.

4.2.3.4 LAMAR2 family

LAMAR2 is a previously undescribed Tc1/mariner element that, like Dumbo, is Afrotherian-specific and is found in both elephant and tenrec. Its 154 amino acid putative transposase has 59% identity with the HSMAR2 transposase. LAMAR2 spawned 1 MITE family, LAMAR2_NA_1_Et.

LAMAR2 was first discovered in elephant based upon its similarity to HSMAR2 and this species-specific variant was named LAMAR2_La. After masking the tenrec genome with the LAMAR2_La consensus, it was found that LAMAR2 was also present in tenrec and the consensus for tenrec, LAMAR2_Et, was constructed. The elephant and tenrec consensuses have 95% nucleic acid identity over the entire sequence and 91% identify at the amino acid level. RepeatMasker did not find LAMAR2 in any other mammalian genomes.

CSA analysis revealed lineage-specific activity of LAMAR2_La, as PIES could be identified in tenrec (Figure 4.23). This finding was corroborated by the low average divergence of LAMAR2_La (11.6%) and the calculated age of 45 myr. Because of the low copy number (n=638) (Figure 4.24), LAMAR2_La did not meet the TCF connectedness criteria and was not ranked.

```
LAMAR2_La site in tenrec
Elephant position:   scaffold_10689:55424-56351
Tenrec position:     scaffold_290110: 975-1175
Elephant:            AAACTATAAATACGACGGTAAG...CTTACCCTCATACAGTATCAT
Tenrec:              aaagtgcaaata----------...----------cagttccat
```

Figure 4.23 Pre-insertion empty sites in tenrec for elephant LAMAR2_La element. Target site duplications are in bold and underlined.

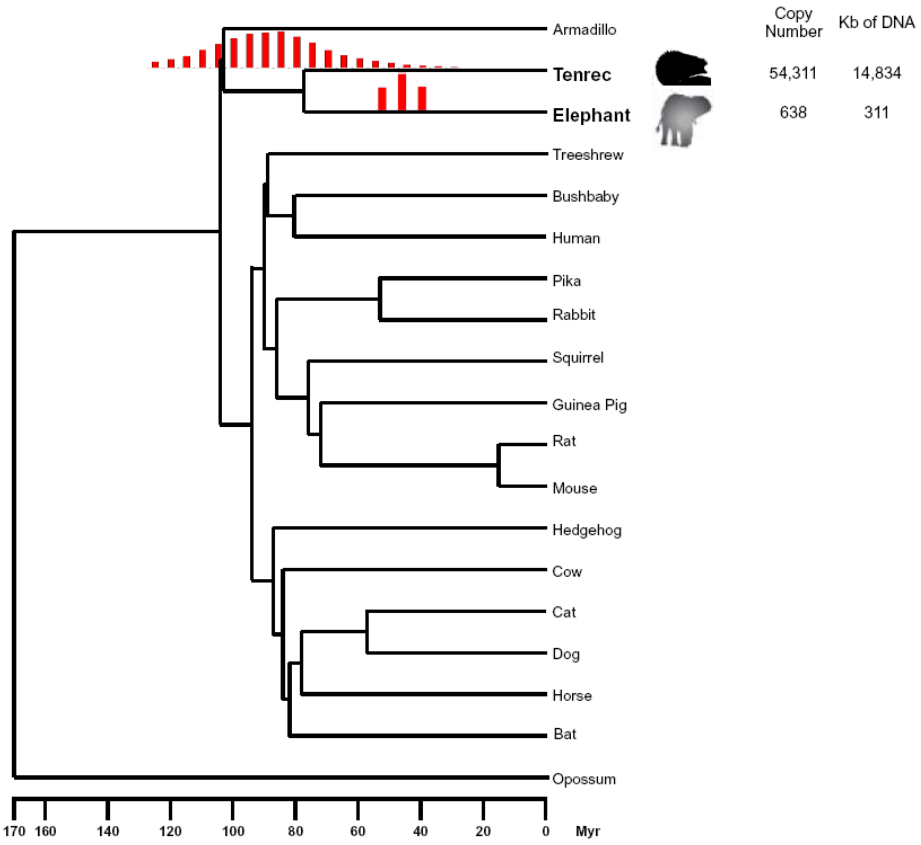


Figure 4.24 Species distribution and timing of amplification of LAMAR2 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring LAMAR2 tribe transposons are in bold. The timing of LAMAR2 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all LAMAR2 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

Conversely, the age of LAMAR2_Et was calculated to be 90 myr. However, CSA revealed no orthologous copies of LAMAR2_Et or its corresponding MITE family LAMAR2_NA_1_Et in elephant. Instead, PIES could be identified in elephant (Figure 4.25). As such, we must conclude that LAMAR2 was not present in the common ancestor of tenrec and elephant, even though the age of LAMAR2_Et suggests this.

```
LAMAR2_NA_1_Et site in elephant
Tenrec position:     scaffold_289728:48317-49462
Elephant position:   scaffold_54542:4811 -5226
Tenrec:              GTAGTCAAGTAAGAAGGTAAGT...ACTTATCCGGGTATACATTTAA
Elephant:            GTAGTCAAGTAC----------...------------ATA--TTTAA
```

Figure 4.25 Pre-insertion empty site in elephant for tenrec LAMAR2_NA_1_Et element. Target site duplications are in bold and underlined.

The discrepancy between the ages of LAMAR2_Et and LAMAR2_La is puzzling. However, this is similar to the discrepancy noted for the Dumbo family of elements, which was present in the common ancestor of elephant and tenrec. Given that the age of LAMAR2_Et in tenrec slightly predates the divergence of tenrec from elephant, we may conclude that our substitution rate for tenrec during this period is calibrated to quickly. However, a second possible explanation for these data may be equally as likely. Since the invasions into these species were independent, we cannot rule out the hypothesis that LAMAR2 invaded tenrec before it invaded elephant and, thus, the timings of the LAMAR2 transfers did not coincide with each other. Therefore, though we cannot conclusively explain the difference in age of LAMAR between the two species, we can state that the infiltrations were indeed independent and produced lineage-specific activity in both species (Figure 4.25).

4.2.3.5 OAMAR family

Oamar1 is an autonomous Tc1/mariner element that has been annotated as Cetartiodactyla-specific in Repbase Update. The transposase of Oamar1 has 45% identity with HSMAR1 at the amino acid level. Masking of the mammalian genomes with the Oamar1 consensus revealed that no copies were found in any other species. The average divergence of Oamar1 copies to their consensus sequence is 14.7%, similar to that of HSMAR1_Bt (14.0%) and Charlie3_Bt (16.7%). This divergence results in a calculated age of 53 myr, well after the divergence of cow from the other Laurasiatherian mammals (84 mya). Oamar1 produced 1 MITE, Oamar1_NA_1, which is a previously unannotated family. This 80 bp MITE achieved a copy number of over 8,000. Dating of Oamar1_NA_1 based on sequence divergence placed

the age at 63 myr. Together, these two families of elements produced 1.1 Mb of DNA in the cow genome.

Next, NIA was used to corroborate the calculated date. The youngest TE Oamar1 nested within was Tigger1_Art, a 72 myr old DNA TE family, while Oamar1_NA_1 nested within Oamar1. TCF ranked Oamar1_NA_1 as approximately the same age as the Charlie3 descended MITEs MER1A_Bt_1 and MER1A_Bt_2 (77 and 68 myr, respectively). These data provides further support for the calculated ages of Oamar1 and Oamar1_NA_1.

Since RepeatMasker found no copies of Oamar1 in any other species and the dates of the Oamar1 tribe families postdate the divergence of cow from the Laurasiatherian ancestor, we had evidence to believe that the presence of Oamar1 in cow was due to an independent horizontal transfer event that occurred strictly in the cow lineage. To investigate this possibility, we used CSA to verify the absence of orthologs in closely related species and to identify PIES. As expected, no orthologous copies of either Oamar1 or Oamar1_NA_1 could be found in dog, but PIES could be identified (Figure 4.26). Therefore, we can conclude that the horizontal transfer of the Oamar1 tribe of elements into cow was an independent event that occurred approximately 50 to 60 mya, at or near the time of the infiltration of HSMAR1 and Charlie3 (Figure 4.27).

```
Oamar1_NA_1 site in dog
Cow position:    chr1:4459428-4459906
Dog position:    chr31:27629317-27629714
Cow:             AATACCAAAATACTAGGTTGGC...GTCAACTCAATGTGAATTTATATTATTCGAA
Dog:             AATGCCAAAATA----------...------------TGAATAAGTATTTTTCCAA
```

Figure 4.26 Pre-insertion empty site in dog for cow Oamar1_NA_1 element. Target site duplications are in bold and underlined.

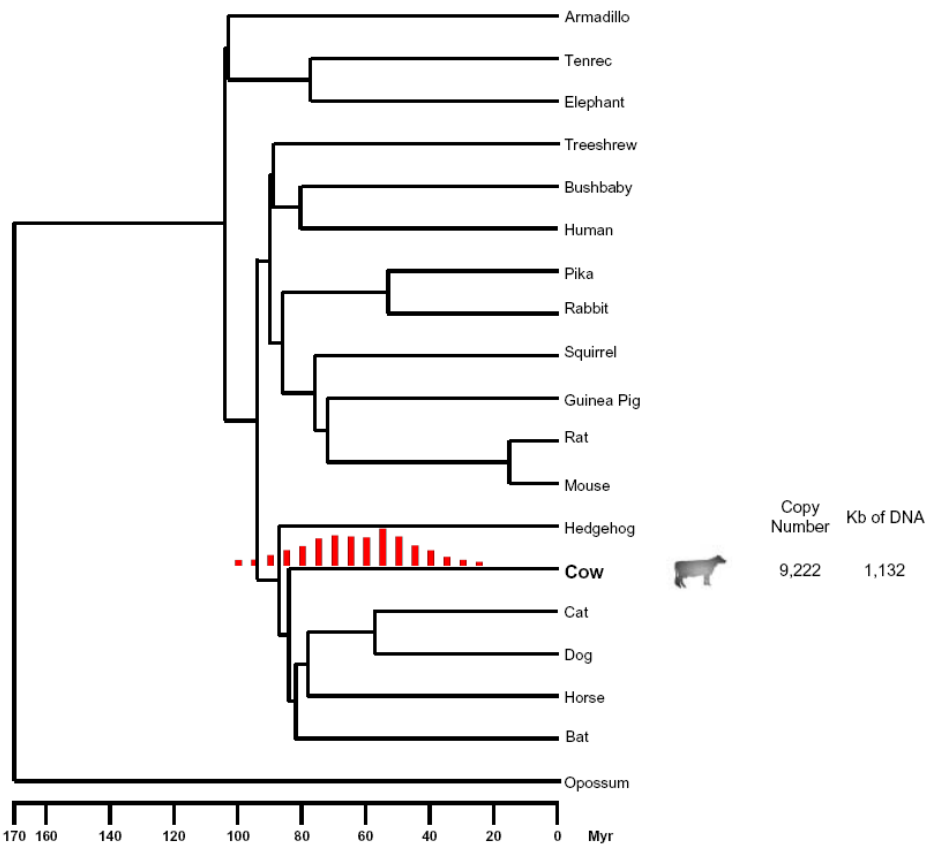


Figure 4.27 Species distribution and timing of amplification of OAMAR1 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring OAMAR1 tribe transposons are in bold. The timing of OAMAR1 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all OAMAR1 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

4.2.3.6 OGMAR1 and OGMAR2 families

OGMAR1 and OGMAR2 are previously undescribed prosimian-specific autonomous Tc1/mariner elements. OGMAR1 has 63% identity with HSMAR1 at the amino acid level over the entire length of the transposase ORF (206 aa). The OGMAR2 transposase (252 aa) has 35% identity with HSMAR2 over the 160 N-terminal amino acids and 31% identity with the Afrotherian-specific LAMAR2 element over the 116 N-terminal amino acids. Unlike OGMAR1, OGMAR2 produced 2 distinct MITE families, OGMAR2_NA_1 and OGMAR2_NA_2. No copies

of OGMAR1 or OGMAR2 and its corresponding MITEs were identified in any other mammalian species when their genomes were masked with the consensus sequences. However, using the consensus sequences as queries for Blastn searches allowed us to identify copies in other prosimian species such as *Lemur catta* (ring-tailed lemur), *Eulemur macaco macaco* (blue-eyed black lemur), *Microcebus murinus* (gray mouse lemur), *Propithecus verreauxi* (Verreaux's sifaka), *Galago senegalensis* (Senegal galago) and *Nycticebus coucang* (slow loris). Given that bushbaby most likely diverged from other prosimians approximately 57 mya (Steiper and Young 2006), OGMAR1 and OGMAR2 must have invaded the prosimian lineage shortly after the divergence of prosimians from the primate lineage.

In order to verify the RepeatMasker results that OGMAR1 and OGMAR2 were not present in other primate species, we searched for orthologous copies of each family in other primates. No copies could be found and PIES could be identified (Figure 4.28a-b). Therefore, based upon CSA, we were able to conclude that OGMAR1 and OGMAR2 are indeed prosimian-specific.

```
Ogmar1 site in human
Bushbaby position:    scaffold_1755.1-117770:108264-109916
Human position:       chr9:32235674-32236060
Bushbaby:             GTCTAAATACTAGATTATTAAG...AATGACCTAGTATAATTTGC-CA
Human:                GGCCAGATGTGA----------...----------TAAGTTGGTCA
```

(a)

```
Ogmar2 site in human
Bushbaby position:    scaffold_101479.1-398088:116975-118271
Human position:       chr2:226010057-226010441
Bushbaby:             ACTTGATATATAAGCTCTGACA...AGACTTTATATAATGATTGTAA
Human:                ACTCAGTATATG----------...----------ACAGTTGTGA
```

(b)

Figure 4.28 Pre-insertion empty sites in (a) human for bushbaby OGMAR1 and (b) human for bushbaby OGMAR2 element. Target site duplications are in bold and underlined.

The OGMAR1 and OGMAR2 tribes are between 17.1% and 18.7% diverged from their respective consensus sequences. As a result, the calculated ages range from 58 to 63 myr old. This date places the infiltration time of both elements just after the divergence of prosimians

from anthropoid primates and just before the divergence of bushbaby from other prosimians (Figures 4.29 and 4.30). This dating is in excellent agreement with the CSA data. NIA further corroborates the dating of OGMAR1 and OGMAR2. According to the TCF rankings, both families are older than the SPIN and OposCharlie1 tribes of elements (both ~30 myr) but near the same age as the Charlie12 (78 myr) and Charlie3 (66 myr) tribes of elements.

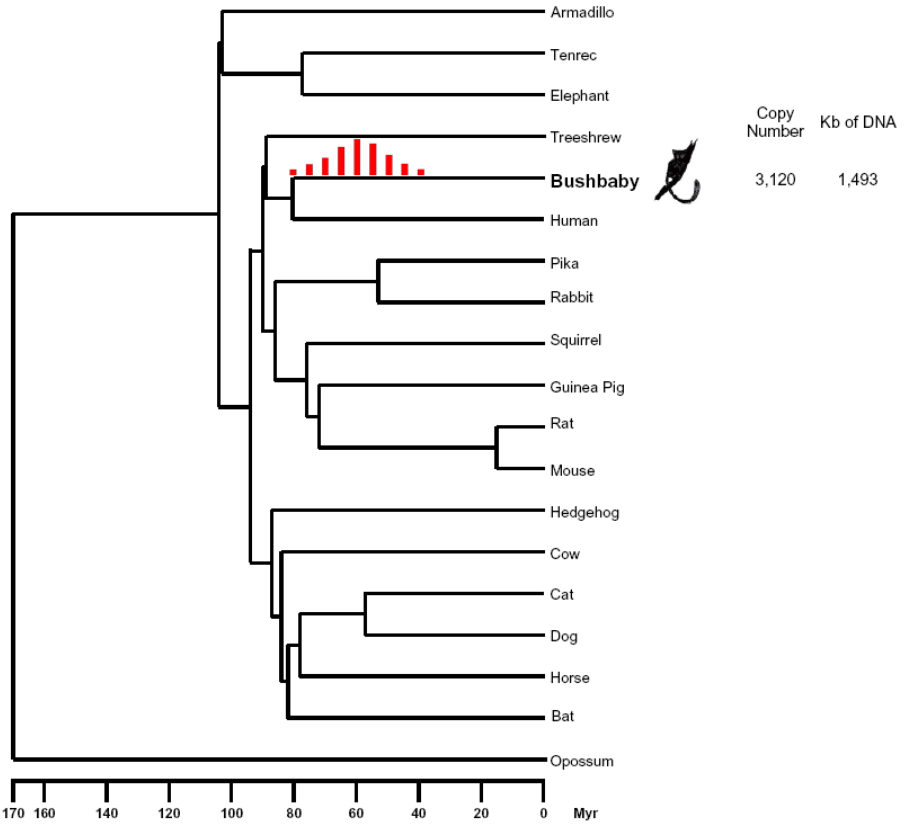


Figure 4.29 Species distribution and timing of amplification of OGMAR1 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring OGMAR1 tribe transposons are in bold. The timing of OGMAR1 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all OGMAR1 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

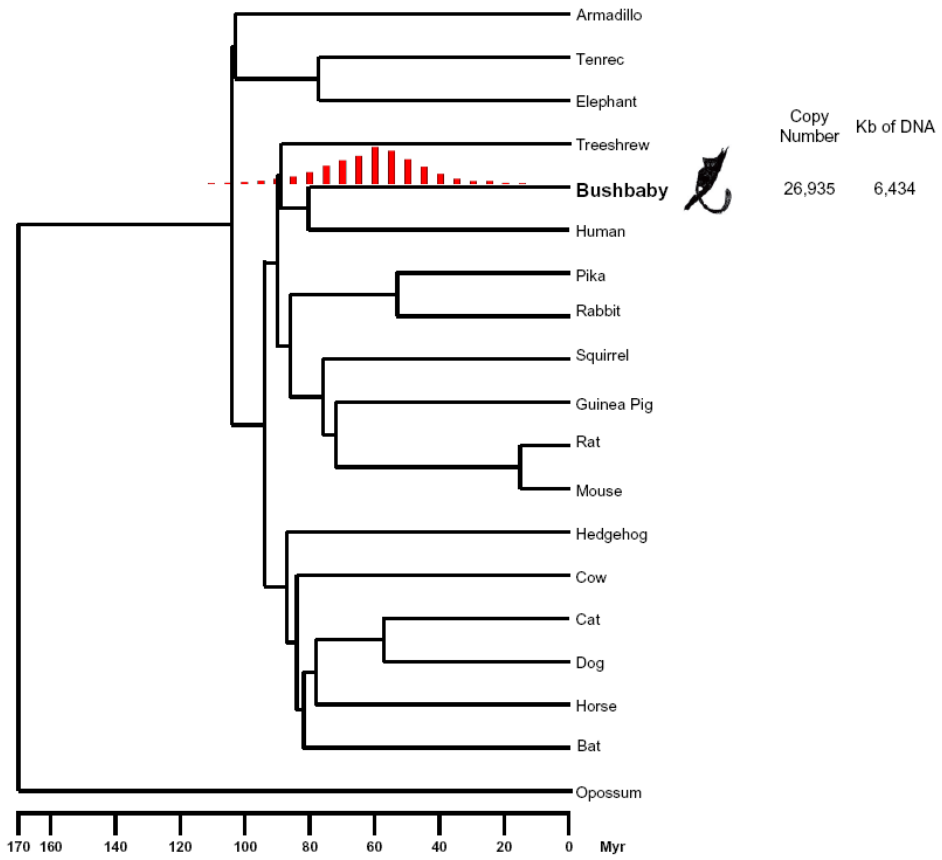


Figure 4.30 Species distribution and timing of amplification of OGMAR2 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring OGMAR2 tribe transposons are in bold. The timing of OGMAR2 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all OGMAR2 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

OGMAR2 was significantly more active than OGMAR1. While OGMAR1 only reached a copy number of 3,120, OGMAR2 and its MITE families achieved a copy number of almost 27,000. Given that both OGMAR1 and OGMAR2 may have been transferred into the prosimian lineage by the same vector, the difference in their transpositional success is intriguing.

4.2.4 Horizontal transfer of piggyBac elements

4.2.4.1 MER85 family

MER85 is a non-autonomous piggyBac element that was previously annotated as anthropoid-specific (Pace and Feschotte 2007). This element appears to have been mobilized by an autonomous element, termed MER85_Auto for simplicity, that was domesticated and is now a part of the PGBD3 gene in human. Three near full-length copies of the autonomous element can be identified in human, along with several other copies that are approximately half-length. MER85 appears to be a deletion derivative of this element, assuming that MER85_Auto was highly similar in sequence to PGBD3.

In our previous analysis, we used CSA to demonstrate that MER85 was indeed anthropoid-specific. We found that MER85 was approximately 40 myr old, that it was nested within primate-specific Alu and LINE elements and that PIES could be identified in bushbaby (Pace and Feschotte 2007). However, masking of the other mammalian genomes revealed that 186 copies of an element similar to MER85 were present in bushbaby. We reconstructed the consensus sequence for this element and found that it was 94 bp longer than MER85 (234 and 140 bp, respectively), and had approximately 90% nucleotide identity over the entire length of MER85. We named this element MER85_Og_1. Alignment of MER85_Og_1 with PGBD3 showed that, like MER85, it was a clear deletion derivative of the putative autonomous element (Appendix E).

We next searched for the presence of MER85_Auto by using the PGBD3 sequence as a query in Blastn searches against bushbaby and were able to find multiple copies. However, using CSA, we found that no copies were orthologous with those in human.

Just as PIES for the human MER85 family had been identified in bushbaby, we were able to find PIES for bushbaby-specific MER85_Og_1 elements in human (Figure 4.31). In addition, no copies of the human MER85 were found at orthologous loci in bushbaby, as no copies of MER85_Og_1 were found at orthologous loci in human. These data clearly show

lineage-specific insertions for MER85_Og_1 and MER85. Since few copies of MER85_Og_1 are present, it did not meet the connectedness criteria for TCF.

```
MER85_Og_1 site in human
Bushbaby position:      scaffold_102854.1-53847:37116-37354
Human position:         chr12:14322972-14323154
Bushbaby:               TGAAGGACTATTAACCCACATATAC...cacagatggGTTAAGAAGC
Human:                  TGAAGGACTATGAA----------...--------------GAAGT
```

Figure 4.31 Pre-insertion empty sites in human for bushbaby MER85_Og_1 element. Target site duplications are in bold and underlined.

We next derived the age of MER85_Og_1 to discover if it corroborated the CSA data. Indeed, the average divergence of MER85_Og_1 is 13.9%, in good agreement with the divergence of MER85 (7.7%) considering that the estimated neutral substitution rate in bushbaby is 2.96x10-9 as compared to 2.06x10-9 in human. Thus, the estimated age of MER85 is 37 years while that of MER85_Og_1 is 47, both of which postdate the divergence of these species. The ages of these elements, in conjunction with the CSA data, reveal that MER85_Og_1 is indeed a bushbaby-specific element.

While neither MER85 or MER85_Og_1 were duplicated to high copy number (923 and 186, respectively), the autonomous element responsible for their transposition appears to have invaded 2 different branches of the primate lineage within the past 50 myr and produced lineage-specific MITEs (Figure 4.32). Interestingly, Newman et al. found that a copy of the autonomous MER85 parent that inserted within an intron of the human ERCC6 gene, a gene involved in DNA repair, formed a fusion transcript with a downstream exon (Newman et al. 2008). This element was domesticated after the divergence of prosimians from the primate lineage and has since been under purifying selection, possibly serving to repress activity of MER85 and its autonomous parent. However, this same insertion did not occur in bushbaby and a PIES could be identified (Figure 4.33). Thus, while the MER85_Auto invaded two different primate lineages, its fate was dramatically different in each.

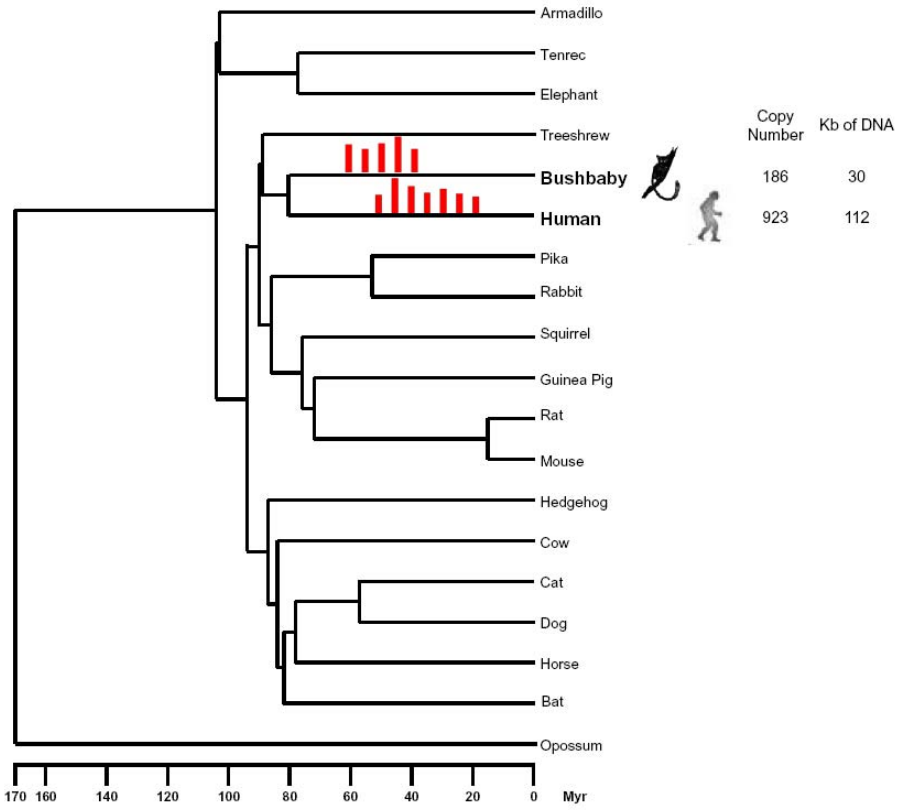


Figure 4.32 Species distribution and timing of amplification of MER85 tribe transposons. The tree depicts the phylogenetic relationship and divergence times of the mammalian species with complete or nearly complete genome sequences currently available. The species harboring MER85 tribe transposons are in bold. The timing of MER85 tribe element amplification in each species lineage is shown above the corresponding branches, by the red vertical bars. Each set of bars represents the age span for all MER85 tribe transposons found in the species, with each individual bar showing the relative proportion of elements falling within the same, non-overlapping 3-myr bin.

```
MER85/PGBD3 site in bushbaby
Human position:     chr10:50392740-50395812
Bushbaby position:  scaffold_103134.1-170382:50512-50989
Human:              GAAAACATCTTTAAcccatttatg...taaatagGTTTTAAAAGTGTAA
Bushbaby:           GAAAACATCCTTAA----------...--------------GGGAGtaa
```

Figure 4.33 Pre-insertion empty sites in bushbaby for human MER85 element located within the PGBD3 gene. Target site duplications are in bold and underlined.

4.3 Discussion

4.3.1 Widespread horizontal transfer of DNA transposons in mammals

We identified 12 families of DNA transposons that, in addition to SPIN, have been independently horizontally transferred into at least one mammalian species or lineage. Four of these families (Dumbo, LAMAR2, OGMAR1, and OGMAR2) were previously unidentified. Together, these 12 families created 44 lineage-specific MITE families (Appendix C) and were responsible for the insertion of between 2,300 and 222,000 copies in each species. In addition, the new transposon copies added between 1.3 and 39.2 Mb of nuclear DNA in each genome (Table 4.1).

Table 4.1 Summary of horizontally transferred DNA TE family activity. Autonomous element families are annotated in bold, with their corresponding MITE families listed in normal font below and indented. Orphan MITE families are italicized.

Species	**TE Name**	**Copy Number**	**Mb of DNA added**
Human	**Charlie12**	48	0.01
	MER30	4,102	0.75
	MER30b	331	0.04
	Charlie3	535	0.33
	MER1A	3,411	1.39
	MER1B	5,933	1.60
	HSMAR1	865	0.48
	MADE1	8,003	0.49
	HSMAR2	1,829	1.11
	MER85	923	0.11
	Total	**25,980**	**6.31**
Bushbaby	**Charlie12**	3,603	1.35
	MER30_Og_1	8,141	1.39
	Charlie3_Og	1,892	0.54
	MER1A_Og_1	1,467	0.32
	MER1A_Og_2	1,407	0.14
	MER1A_Og_3	1,280	0.37
	MER1A_Og_4	2,557	0.54
	MER1A_Og_5	2,388	0.53
	OposCharlie1	5,256	0.29
	OposCharlie1_NA_2_Og	15,829	2.73

Table 4.1 - *Continued*

	HSMAR1_Og	475	0.22
	OGMAR1	3,120	1.49
	HSMAR2_Og	1,439	0.60
	OGMAR2	8,848	3.65
	OGMAR2_NA_1	6,234	0.85
	OGMAR2_NA_2	11,853	1.93
	MER85_Og_1	186	0.03
	SPIN_Og	7,077	0.97
	SPIN_NA_1_Og	8,474	1.46
	SPIN_NA_2_Og	17,436	1.32
	Total	**108,962**	**20.75**
Mouse	**Charlie12**	528	0.13
	RMER30	2,851	0.35
	RCHARR1	6,411	1.93
	MMAR1	311	0.10
	SPIN_NA_10_Rode	34,828	5.96
	Total	**44,929**	**8.47**
Guinea Pig	**Charlie12**	299	0.06
	RMER30	483	0.04
	Charlie3	1,808	0.95
	MER1A	1,174	0.19
	MER1B	550	0.09
	Total	**4,314**	**1.33**
Bat	**Charlie12**		
	MER30_MI_1	1,575	0.11
	MER30_MI_2	229	0.03
	MER30_MI_3	3,577	1.32
	OposCharlie1	148	0.03
	OposCharlie1_NA_3_MI	288	0.06
	OposCharlie1_NA_4_MI	415	0.07
	OposCharlie1_NA_5_MI	624	0.11
	OposCharlie1_NA_6_MI	363	0.05
	nhAT2	6,051	1.18
	SPIN_MI	2,735	0.51
	SPIN_NA_7_MI	21,124	4.13
	SPIN_NA_8_MI	3,619	0.63
	SPIN_NA_9_MI	10,578	3.10
	SPIN_NA_10_MI	11,901	2.35

Table 4.1 - *Continued*

	Total	**63,227**	**13.69**
Treeshrew	**Charlie3**	1,219	0.31
	MER1A_Tb_1	831	0.24
	MER1A_Tb_2	792	0.18
	MER1A_Tb_3	1,544	0.37
	MER1A_Tb_4	1,536	0.42
	HSMAR1_Tb	1,897	0.73
	HSMAR2_Tb	9,878	1.81
	HSMAR2_NA_1_Tb	54,342	5.76
	Total	**72,039**	**9.82**
Cow	**Charlie3_Bt**	4,722	0.68
	MER1A_Bt_1	3,986	0.65
	MER1A_Bt_2	2,543	0.35
	HSMAR1_Bt	2,648	1.78
	Oamar1	1,153	0.57
	Oamar1_Na_1	8,069	0.56
	Total	**23,121**	**4.60**
Tenrec	**OposCharlie1**	6,505	0.55
	OposCharlie1_NA_1_Et	7,328	1.40
	RCHARR1_Et	1,223	0.19
	Dumbo	2,362	0.44
	Dumbo_NA_1	1,897	0.35
	Dumbo_NA_2	2,086	0.36
	HSMAR2_Et	5,503	1.85
	HSMAR2_NA_4_Et	42,136	2.95
	LAMAR2_Et	27,458	6.38
	LAMAR2_NA_1_Et	26,853	8.46
	SPIN_Et	13,960	3.58
	SPIN_NA_1_Et	52,541	6.59
	SPIN_NA_6_Et	32,817	6.06
	Total	**222,669**	**39.17**
Elephant	**Dumbo**	2,919	0.73
	Dumbo_NA_1	4,149	1.00
	Dumbo_NA_2	2,762	0.56
	LAMAR2_La	638	0.31
	Total	**10,468**	**2.60**
Armadillo	**HSMAR1_Dn**	2,347	1.50
	Total	**2,347**	**1.50**

Table 4.1 - *Continued*

Pika	**HSMAR2**	7,998	1.40
	HSMAR2_NA_2_Op	46,014	4.48
	Total	**54,012**	**5.88**
Rabbit	**HSMAR2_Oc**	12,097	2.17
	HSMAR2_NA_3_Oc	86,187	8.32
	Total	**98,284**	**10.50**
	Grand Total	**730,352**	**124.63**

Surprisingly, all autonomous TE families that we previously classified as anthropoid-specific (Charlie12, Charlie3 and HSMAR1) (Pace and Feschotte 2007) were found in other eutherian mammals, although the species distributions of each family differ. The 4 families of elements previously classified as rodent-specific (MMAR1, RCHARR1, RMER30 and URR1) (Waterston et al. 2002) were also found in other eutherian species. Therefore, while it could be stated accurately that these families produced lineage-specific activity in the anthropoid primate or rodent lineages, our data now clearly demonstrate that they have independently amplified in other lineages as well (Figure 4.34).

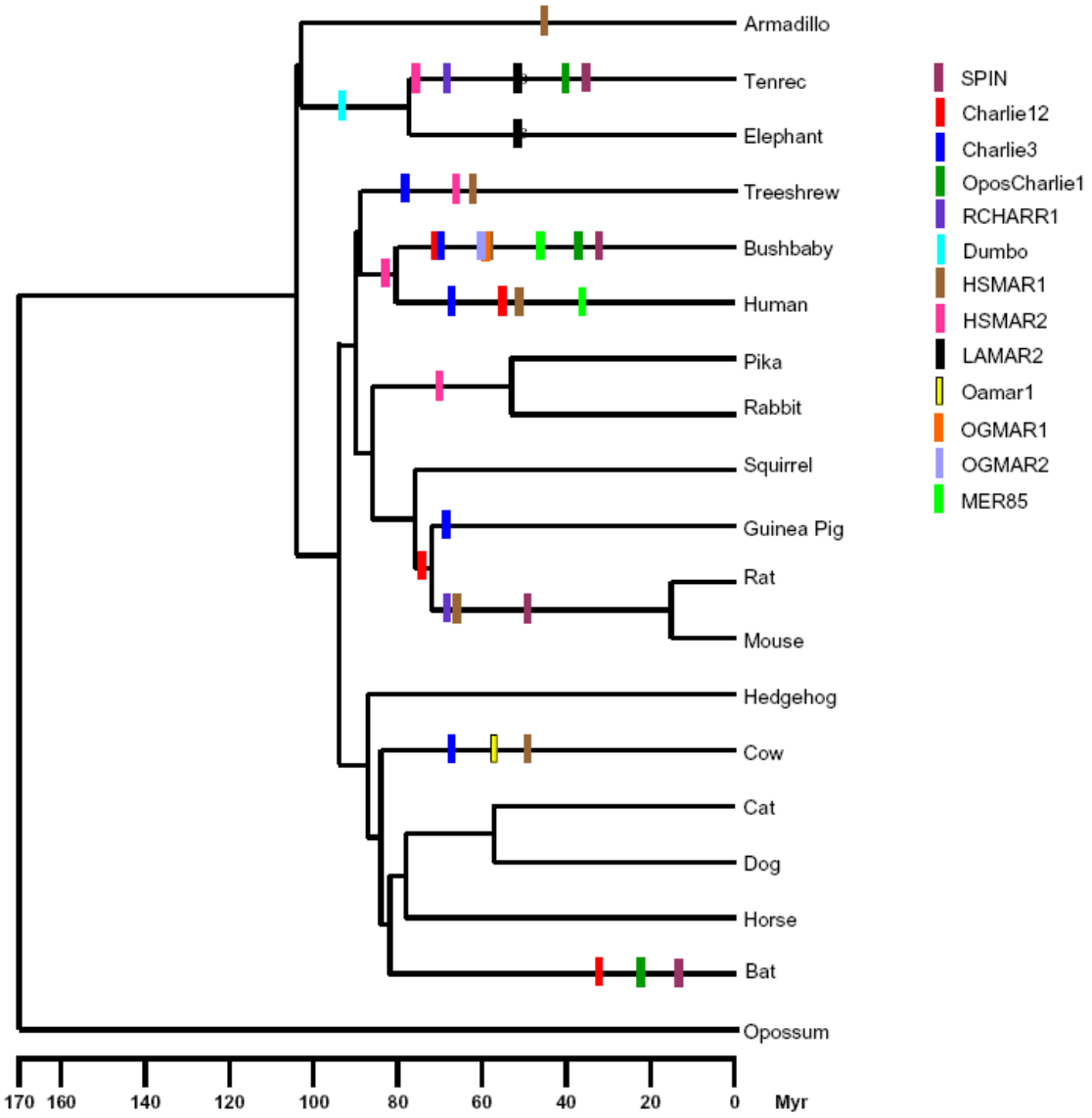


Figure 4.34 Mammalian phylogeny with insertion dates for horizontally transferred DNA TEs. Each rectangle represents a single, independent horizontal transfer event into a species or lineage. The placement of the rectangle is based upon the average age of all elements within the tribe.

Analysis of the eutherian genomes into which a DNA transposon family was horizontally transferred reveals that several species and lineages appear to have been more prone to HT than others in the last 80 myr. In particular, the tenrec and bushbaby seem to be

especially susceptible, as they suffered invasions from 6 and 9 families of autonomous elements, respectively. Conversely, our data suggest that other lineages, such as carnivores (dog and cat), horse, hedgehog and squirrel have been less prone to HT as none of these lineages have been infiltrated by any known DNA transposons in the last 80 myr (Figure 4.34). We must note, however, that we cannot exclude the possibility that as of yet undiscovered TE families may have invaded these species or that these species may have been geographically isolated from the other species that suffered horizontal transfer, thereby preventing their exposure to the HT vector.

We were unable to distinguish any patterns for which species would suffer the HT of a family of elements, as almost all families exhibit a different species distribution (Figure 4.34). One exception to this observation is the OposCharlie1 and *SPIN* families of elements. Our data suggests that both families were potentially co-transferred from the same source since both are found in the same mammalian species (tenrec, bushbaby, bat and opossum) and are approximately the same age in each lineage (Figure 4.34). The only difference between the two is that OposCharlie1 is not found in murine rodents, *X. tropicalis* or *A. carolinensis* while *SPIN* is. Interestingly, the geographic distributions of these species and their ancestors do not appear to overlap, as tenrec was already localized to Madagascar (Poux et al. 2005), opossum to the Americas (Whitaker Jr. and Hamilton 1998) and bushbaby to Africa (Yoder et al. 2003). The vespertilionid bat most likely had a cosmopolitan geographic distribution. Therefore, we cannot rule out the possibility that other families of elements were potentially co-transferred as well, but had differential success in invading the germline of each species. As more mammalian species from a broader range of clades are sequenced, clear patterns might emerge.

4.3.2 Census of DNA transposons in mammals

We were able to identify a total of 249 DNA transposon families that were found in one or more eutherian mammals with a copy number of at least 100 (Appendix C). Of these families, 146 belonged to the hAT (59%), 92 to the Tc1/mariner (37%), 5 to the piggyBac (2%),

5 to the MuDR (2%) and only 1 to the Merlin superfamily. This is almost twice as many DNA transposon families with a copy number of greater than 100 than we identified previously in primates (n=125). The difference in number can be accounted for by two factors. First, during this work, we discovered a total of 63 new families of elements that are found in at least one eutherian species but not in human. Second, 61 families were previously annotated in other mammals but not in human or their copy number in human was reported as less than 100. Remasking the human genome with our database of all mammalian DNA transposons, which is likely more comprehensive than the database used to mask the latest release of the human genome, revealed that these 61 families are indeed present in human with a copy number of greater than 100. As such, the number of DNA transposon families in the human genome now amounts to 186, an increase of 49%.

The majority of the DNA transposons currently recognizable in eutherian mammals predate the mammalian radiation (Appendix C). Of the 249 families, members of 169 families (68%) are unambiguously present at orthologous loci in eutherian mammalian species from multiple clades, indicating they inserted prior to the divergence of eutherian species. This finding is corroborated by our previous work that classified 68% of the families found in primates as eutherian-wide (Pace and Feschotte 2007). However, the possibility of continuous, lineage-specific activity for some of these families, especially those active near the time of the eutherian mammal radiation cannot be excluded. For example, copies of the Tigger1 family of elements are unambiguously present at orthologous loci throughout eutherian mammals, but many lineage specific copies are found as well (Feschotte and Gilbert, unpublished data). The other 80 families were active after the radiation of the major eutherian clades (~100 mya), and the amplification of these elements has contributed a substantial amount of genomic variation along the major branches of the eutherian phylogeny.

4.3.3 Waves of DNA transposon activity in eutherian mammals

Our current data reveals that 3 major waves of transpositional activity have occurred as a result of the HT of DNA transposons into eutherian mammals. The first wave occurred just prior to the eutherian radiation, approximately 100 mya, and produced 100 hAT, 51 Tc1/*mariner*, 1 piggyBac and 5 MuDR families that are now distributed throughout the eutherian phylogeny.

The second wave occurred between 40 and 80 mya when the Charlie12 and Charlie3 hAT families, the HSMAR1 and HSMAR2 Tc1/*mariner* families and the autonomous element families responsible for the transposition of MER85, MER75 and MER107 independently invaded eutherian lineages (Figure 4.35 and 4.36) spawning lineage-specific MITE families. This was the most intense period of Tc1/mariner activity, adding between 47,000 and 98,000 copies of HSMAR2 and its corresponding MITE families to the tenrec, treeshrew, rabbit and pika genomes.

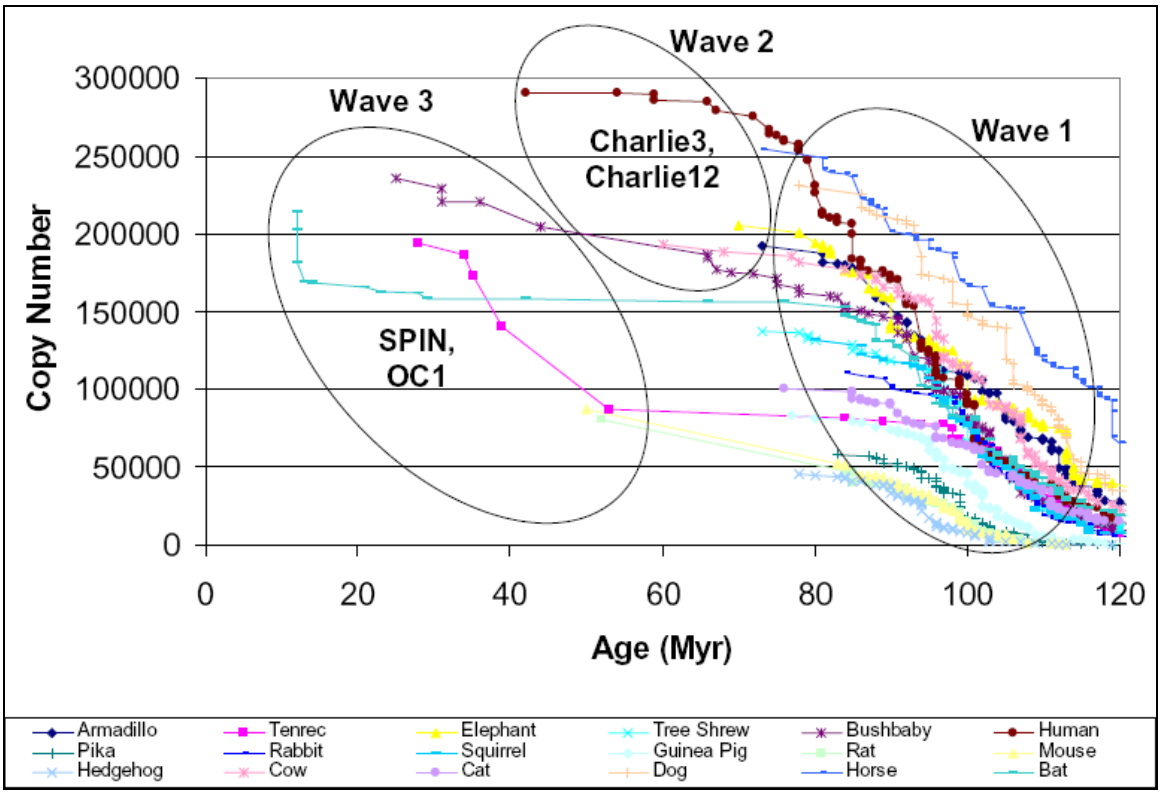


Figure 4.35 Cumulative copy numbers of hAT superfamily transposons in mammalian species.

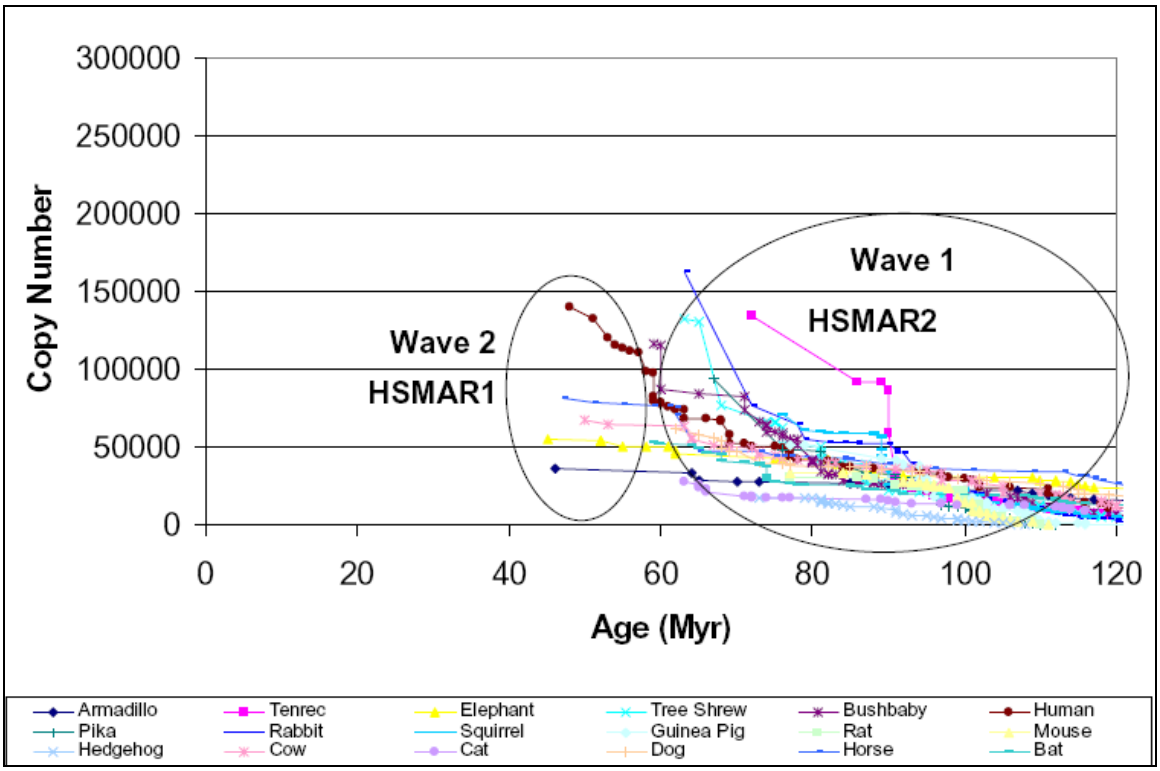


Figure 4.36 Cumulative copy numbers of *mariner* superfamily transposons in mammalian species.

The third, and most recent wave of DNA transposon activity, occurred between 25 and 40 mya as a result of the activity of the OposCharlie1 and *SPIN* hAT superfamily elements (Figure 4.35). This is the only known DNA transposon activity that has occurred in eutherian mammals in the past 40 myr, with the striking exception of the intense activity that appears to be limited to vespertilionid bats (Ray et al. 2008; Ray et al. 2007; Pritham and Feschotte 2007). Interestingly, excluding bats, only hAT elements were transpositionally active during this wave, as no other currently annotated superfamilies of elements appear to have been horizontally transferred during this period. Again, we acknowledge that other, as of yet unidentified families of elements, may have been active during both this and the subsequent waves of activity.

Comparison of the 3 waves of transposon activity in eutherian mammals yields several interesting observations. First, in 5 of the 18 eutherian species (cat, dog, hedgehog, horse and squirrel), the second and third waves were undetectable. Second, the third wave occurred only

in 4 eutherian species: bat, bushbaby, murine rodents and tenrec. These 4 species represent four of the major eutherian mammalian clades (Afrotheria, Laurasiatheria, primates and rodents), supporting the conclusion that infiltrations of DNA transposons and subsequent bursts of activity are not limited to only one clade of eutherian mammals. However, the question can be posed as to why only these species were successfully invaded during the third wave? Were these species located in the same geographic region, isolated from the other species? Did the transposons invade other species but were unable to successfully enter the germline and reach fixation? As new fossil evidence is uncovered and the ability of mammalian genomes to resist infiltrations of transposable elements becomes better understood, these questions may be answered.

As observed previously for primates (Pace and Feschotte 2007; Lander et al. 2001), there has been a general slowdown in the activity of DNA transposons since the eutherian radiation. While separate waves of transposition have occurred during this time, each wave was less successful than the previous in terms of copy numbers and amount of nuclear DNA added. With the exception of bat, we find no evidence for the activity of DNA transposons in the past ~25-30 myr in any of the 18 eutherian mammals surveyed. This raises the perplexing question of why this general extinction occurred?

One possible explanation lies at the Eocene-Oligocene boundary, approximately 35 mya. It has been proposed that during this period a climatic crisis caused a faunal turnover throughout the entire northern hemisphere, resulting in a mass extinction of many Eocene mammals (Kohler and Moya-Sola 1999). Fossil evidence indicates that this extinction resulted in large and medium sized animals being replaced by those with smaller body sizes, especially rodents and lagomorphs (Meng and McKenna 1998). If the vector(s) that was capable of transferring DNA transposons between diverse mammals, such as a DNA virus was harbored primarily in a group of mammals that became extinct during this period, then the lack of recent, widespread HT events may be explained. Additionally, other intermediate vectors such as

blood-feeding invertebrates or parasites that became extinct during this time, may have also succumbed to extinction, thereby halting further horizontal transfers. Assuming that DNA transposons were transferred through a vector species, than a different vector may be responsible for the recent transfer of DNA transposons to bats.

Perhaps the most significant finding of our study is that the accepted idea that DNA transposons cannot be horizontally transferred between, or independently infiltrate multiple mammalian species, is not correct. The 13 families we identified as being horizontally transferred (including *SPIN*) infiltrated 11 of the 18 genomes we surveyed (61%). While we admit that our investigation is somewhat biased because we could only survey those species for which draft genome sequences are publicly available, there is no reason to believe that this sampling is not representative of eutherian mammals as a whole since it includes species from 6 branches of the eutherian clade (Afrotheria, Lagomorpha, Laurasiatheria, primates, rodents and Xenarthra). Indeed, we believe that as more mammalian genomes are sequenced, the widespread horizontal transfer of DNA transposons will become more apparent, perhaps even becoming the accepted paradigm.

4.4 Methods

4.4.1 Annotation of known mammalian DNA transposons

In order to thoroughly investigate the evolutionary history of DNA transposons in mammals, we first had to identify as many DNA TEs as possible. In previous investigations, we noticed that some TEs in Repbase Update, and thus in the RepeatMasker output, could confidently be identified as the same element in different lineages, but were designated by different names, i.e., MMAR1 in mouse and HSMAR1 in human (95% nucleotide identity over the entire length of the elements). The different names for the same element thus made identification of DNA TEs present in more than one mammalian species difficult.

To address this problem, we began by creating a comprehensive list of all annotated DNA TEs found in 18 different placental mammals and 1 marsupial (opossum). The 18

placental mammals encompassed 6 major clades: Xenarthra (armadillo), Afrotheria (tenrec and elephant), primates (treeshrew, bushbaby and human), Rodentia (mouse, rat, guinea pig and squirrel), Lagomorpha (pika and rabbit) and Laurasiatheria (hedgehog, cow, cat, dog, horse and bat), following the classification of Kriegs et al (Kriegs et al. 2006). We then constructed a Blast database of all TEs on the list and performed Blastn searches using each TE as the query sequence against the database. This process allowed us to confidently identify TEs that were annotated with multiple names in Repbase. For annotated TEs with two or more different names and no apparent lineage-specific activity in mammalian species, we chose one name to use in all species. The group of elements the TE was classified in dictated the choice of name. For example, Arthur1A, a non-autonomous element of the Arthur1 family, was also annotated as MER69A. In order to avoid confusion, all instances of MER69A were reannotated as Arthur1A. For those TEs that were annotated with different names, were highly similar (>= 95% nucleotide identity) and that appeared to have lineage-specific activity, the name of the TE was not changed, as with MMAR1 and HSMAR1, to indicate the lineage-specific activity. This is the same procedure that was used previously to denote lineage-specific SPIN elements (Pace et al. 2008).

In addition to a lack of consistent naming of TEs between species, there also was no comprehensive classification of all mammalian DNA transposons into tribes. For instance, HSMAR1 and its corresponding MITE, MADE1, a deletion derivative of HSMAR1 that most likely mobilized by HSMAR1, would be classified together as the HSMAR1 tribe. Again using each TE as a query sequence in Blastn searches against the database of annotated DNA TEs, we classified all mammalian TEs into tribes based upon structural similarities such as terminal inverted repeats (TIRs) and nucleotide identity between autonomous and non-autonomous elements (Appendix C). This classification not only allowed us to determine which TEs were structurally similar, but also to identify the putative autonomous TEs responsible for the mobilization of their corresponding non-autonomous MITE families.

After the comprehensive list of DNA TEs was assembled, all 19 genomes were remasked with RepeatMasker using a database composed of the consensus sequences for these TEs. The resulting output was then analyzed to determine the presence or absence of DNA TEs within each species (for examples, see Figure 4.37). As expected, many TEs that were previously unannotated within certain species were indeed present, such as Charlie3 and HSMAR1. This masking also provided an initial annotation of DNA TEs for species where no public TE annotation was currently available (tree shrew, bushbaby, pika, squirrel, guinea pig, hedgehog and bat).

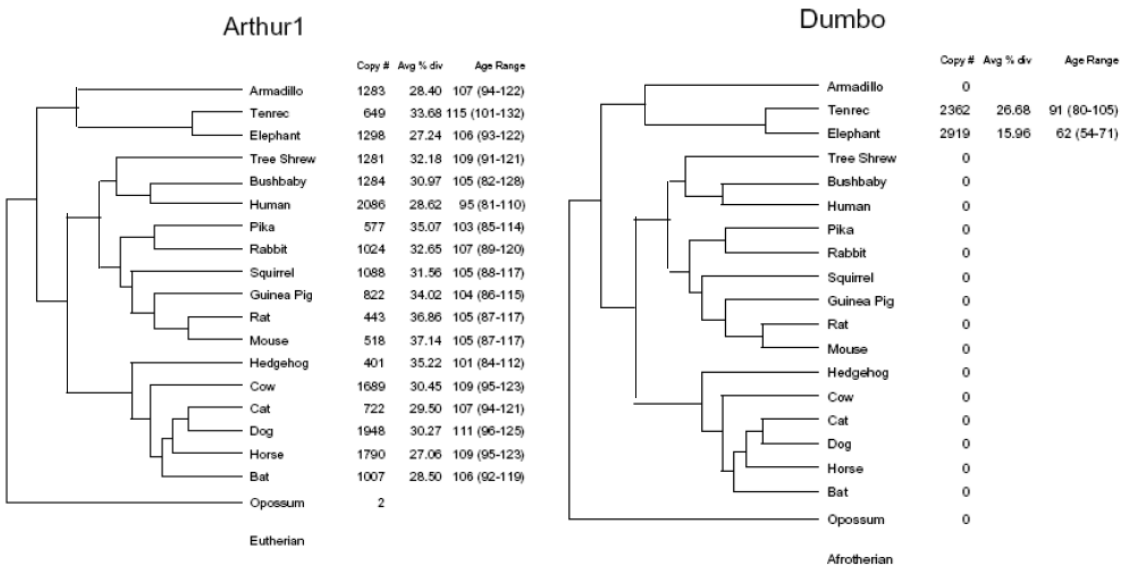


Figure 4.37 Two examples of mammalian phylogenies showing RepeatMasker results

4.4.2 Construction of consensus sequences

Once one copy of a putative TE was identified, this sequence was used as the query sequence in a Blastn search against the genome in which the element was found. When possible, the sequences of the top 20 HSPs that matched the entire length of the query sequence, along with 20 bp flanking either side, were extracted and aligned using ClustalX (Thompson et al. 1997) from within BioEdit (Hall 1999). The consensus was then constructed

using the majority rule method. If one nucleotide was not present in the majority of the sequences at a certain position, the nucleotide that was present in the most sequences was used in the consensus.

4.4.3 Calculation of the age of TE families

The age of TE families was calculated using the method previously employed to date SPIN families (Pace et al. 2008). All ages were corrected with the Jukes-Cantor correction for multiple substitutions. The neutral substitution rates that were used are in Table 4.2.

Table 4.2 Neutral substitution rates used to calculate age of TE families. Substitution rates are in number of substitutions per million years.

Group	DNA			SINE			LINE			LTR			Date of divergence from human (myr)		
	Average Rate	High Rate	Low Rate	Average Rate	High Rate	Low Rate	Average Rate	High Rate	Low Rate	Average Rate	High Rate	Low Rate	Median	Low	High
Armadillo	2.6563	3.0358	2.3215	2.4703	2.8232	2.1589	2.6128	2.9861	2.2835	2.8676	3.2772	2.5061	104	91	119
Bat	2.6920	3.0859	2.3872	2.5896	2.9685	2.2964	2.8281	3.2420	2.5080	3.1438	3.6038	2.7879	94	82	106
Bushbaby	2.9590	3.7575	2.4155	2.8388	3.6048	2.3174	3.1398	3.9871	2.5631	3.2951	4.1843	2.6899	80	63	98
Cat	2.7457	3.1475	2.4349	2.7947	3.2037	2.4784	2.7666	3.1715	2.4534	3.2136	3.6839	2.8498	94	82	106
Cow	2.8000	3.2098	2.4831	2.8250	3.2384	2.5052	2.8358	3.2508	2.5148	3.1027	3.5568	2.7515	94	82	106
Dog	2.7369	3.1375	2.4271	2.8389	3.2544	2.5176	2.7892	3.1974	2.4735	3.1031	3.5572	2.7518	94	82	106
Elephant	2.5652	2.9317	2.2419	2.3572	2.6939	2.0601	2.4901	2.8458	2.1762	2.7809	3.1782	2.4304	104	91	119
Guinea pig	3.2796	3.9444	2.9483	3.1551	3.7946	2.8364	3.3178	3.9903	2.9827	3.6562	4.3973	3.2869	89	74	99
Hedgehog	3.4996	4.2090	3.1461	3.2819	3.9471	2.9504	3.1070	3.7368	2.7932	3.7208	4.4750	3.3450	89	74	99
Horse	2.4804	2.8433	2.1996	2.4364	2.7929	2.1606	2.5502	2.9234	2.2615	2.8481	3.2649	2.5257	94	82	106
Mouse	3.5411	4.2589	3.1834	3.4400	4.1373	3.0925	3.4309	4.1264	3.0844	3.8513	4.6320	3.4623	89	74	99
Opossum	3.2113	3.4375	2.5964	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	76	71	94
Pika	3.4214	4.1149	3.0758	3.3807	4.0660	3.0392	3.3998	4.0890	3.0564	3.7777	4.5435	3.3961	89	74	99
Rabbit	3.0369	3.6525	2.7302	3.1262	3.7599	2.8104	3.1671	3.8091	2.8472	3.4486	4.1477	3.1003	89	74	99
Squirrel	2.9928	3.5995	2.6905	2.8925	3.4788	2.6003	3.0409	3.6572	2.7337	3.3423	4.0198	3.0047	89	74	99
Tenrec	2.9173	3.3341	2.5496	2.8482	3.2551	2.4892	2.7308	3.1209	2.3865	3.2437	3.7071	2.8349	104	91	119
Treeshrew	2.9551	3.5541	2.6566	2.8954	3.4823	2.6030	3.0066	3.6161	2.7029	3.2537	3.9132	2.9250	89	74	99

4.4.4 Identification of orthologous TEs between species

Orthologous TEs between reference and comparison genomes were identified by using the OrthoBlast program (Pace and Feschotte, unpublished). For each TE in the reference genome that had not nested within another transposon, the TE, along with 200 bp flanking each side, were used as a query sequence in OrthoBlast against the comparison genome. Each putative ortholog identified by OrthoBlast was manually inspected. If more than one copy of the entire query sequence was found, then the TE was determined to not have an ortholog in the comparison genome since the true ortholog could be confidently identified.

APPENDIX A

AGES OF HUMAN LINE, SINE AND DNA TRANSPOSONS

Categories	
1	< 40 My
2	40 - 63 My - Anthropoid-Specific
3	64 - 80 My - Primate-Specific
4	> 80 My - Eutherian-Specific

Category	RepName	Corrected Age (My)	% Div (excluding CpGs)*
1	L1HS	10	0.9
1	L1PA2	13	1.1
1	*AluY*	16	3.3
1	L1PA3	19	1.8
1	*AluSc*	22	5.0
1	*AluSg1*	23	4.3
1	*AluSg*	23	5.7
1	L1PA4	25	2.5
1	*AluSq*	25	6.6
1	*AluSx*	28	7.9
1	L1PA5	31	3.3
2	MER85	37	4.6
1	L1PA6	40	4.9
2	L1PA8	42	6.0
2	*AluJb*	42	17.3
2	MER107	43	6.0
2	MER75B	44	6.6
2	*AluJo*	44	14.7
1	L1PA7	45	5.9
2	MADE1	46	7.5
2	MER30	46	7.0
2	HSMAR1	48	7.9
2	MER75	48	8.5
3	L1MA2	49	19.1
2	L1PA8A	50	10.6
3	L1MA3	51	21.7
2	L1PA10	51	11.2
2	Charlie3	51	8.7
3	L1PA15	53	16.2
3	L1PA16	53	17.4
2	MER1A	53	9.3
2	L1PB1	54	10.1
2	L1PA11	55	11.1
3	L1PB3	57	15.8
3	HSMAR2	57	14.8
3	Tigger1	59	14.8
2	MER30B	60	7.7
2	L1PA14	60	10.9
3	Tigger3b	61	10.9

3	Tigger2	61	16.7
2	MER1B	61	11.8
3	Tigger4(Zombi)	62	10.0
3	MER46B	63	17.0
2	L1PA13	63	8.6
2	L1MA1	63	11.9
3	MER44A	64	13.0
2	L1PB2	64	7.5
3	Tigger2a	64	16.7
3	Tigger3(Golem)	64	10.7
4	MER53	65	21.2
3	Tigger5a	65	16.1
3	Ricksha_c	66	18.8
3	MER46A	66	15.0
3	MER6	66	14.5
3	MER6A	67	15.2
3	MER44C	68	15.0
3	Tigger7	69	12.1
2	L1PA12	70	11.2
4	Tigger6b	70	21.7
3	Ricksha_b	70	29.7
3	Tigger5	70	19.0
4	MER96B	70	20.1
3	MER44D	70	18.0
3	Ricksha	71	13.4
3	MER6C	71	22.2
3	MER2B	71	18.5
4	Charlie5	72	23.2
3	MER44B	72	17.0
4	MER58D	72	26.1
4	MER33	72	25.0
4	Tigger6	73	25.9
4	MER45	73	21.0
3	Tigger5b	73	18.2
4	MER3	73	24.5
4	Tigger6a	73	26.7
4	MER45A	73	21.2
4	MER58	74	24.2
4	MER58B	74	18.3
3	MER2	75	22.9
4	MER45B	75	26.1
4	Cheshire	75	30.3
4	MER58A	75	23.1
3	MER8	76	17.2
4	MER45R	76	26.6
3	MER6B	76	18.9
3	MER82	76	19.0

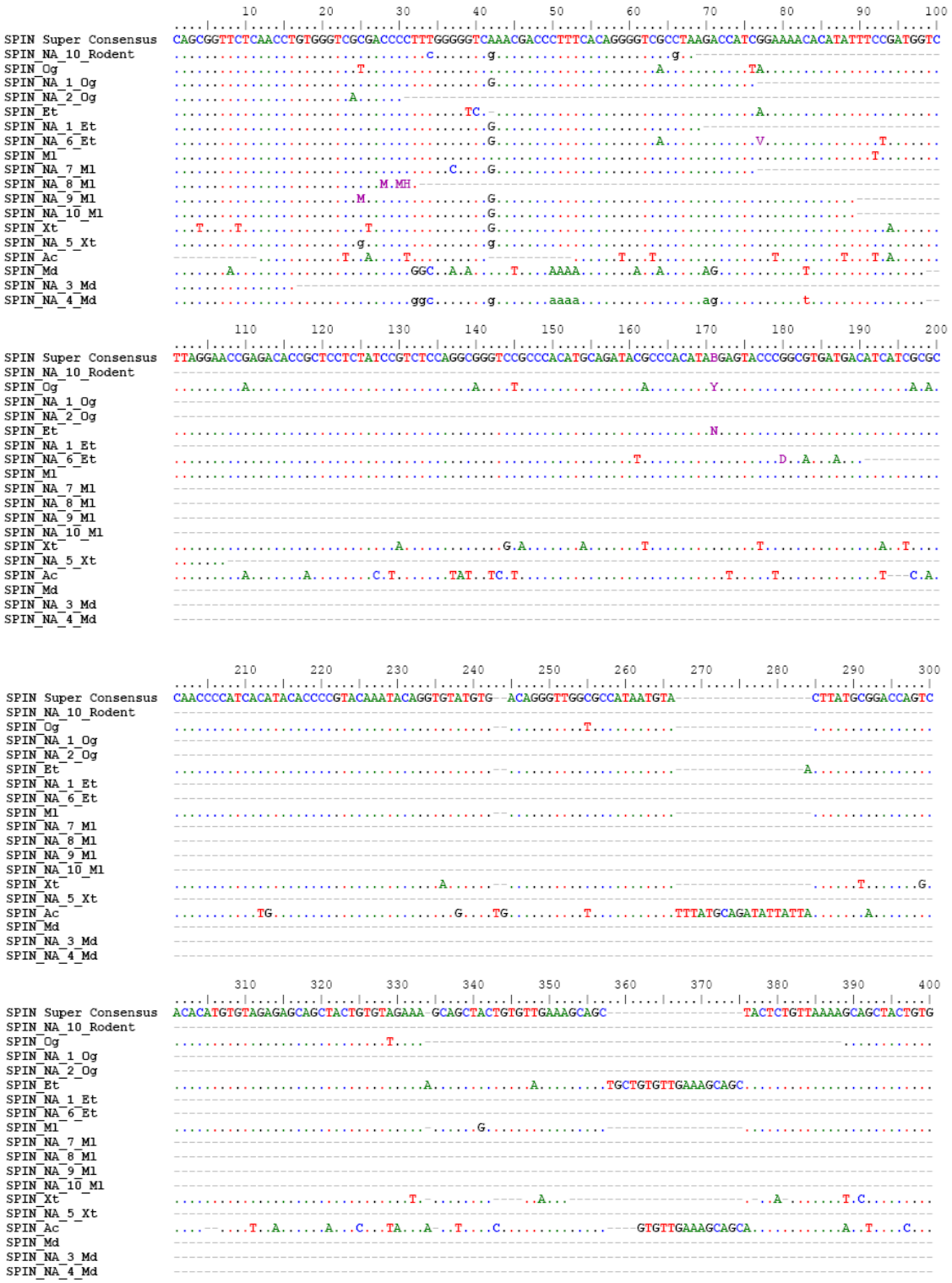
4	Charlie1	77	29.3
4	Looper	77	31.3
4	ORSL	77	25.9
4	Charlie1b	77	30.9
4	MADE2	77	23.3
4	MER20	77	15.5
4	MER63B	77	26.4
3	Tigger5c	78	21.6
4	Charlie1a	78	29.1
4	MER63D	78	26.6
4	MER119	79	29.2
4	MER63C	80	29.7
4	MER106B	80	29.4
4	MER106A	82	21.9
4	MER96	83	23.9
4	MER105	83	20.3
4	MER99	83	24.0
4	MER58C	83	25.9
4	MER63	83	30.0
4	MER97a	84	19.7
4	MER97b	84	32.1
4	Charlie4a	85	33.0
4	MER91	85	38.3
4	MER5A1	86	20.4
4	MER5C	86	33.5
4	Charlie6	86	34.2
4	Charlie4	87	33.3
4	MER97c	87	24.7
4	Charlie9	87	39.8
4	MER104	87	26.6
4	MER81	87	25.5
4	Zaphod2	87	20.7
4	Charlie10	88	35.8
4	Zaphod	88	34.3
4	Arthur1	89	37.4
4	MER69A	90	31.8
4	MER63A	90	43.6
4	Kanga1	92	39.2
4	Kanga1c	92	30.0
4	MER112	92	39.8
4	Charlie2a	92	39.1
4	BLACKJACK	92	38.9
4	Charlie2	93	35.0
4	FordPrefect	94	42.4
4	MER45C	94	20.5
4	MER69B	94	34.9
4	MER94	94	38.5

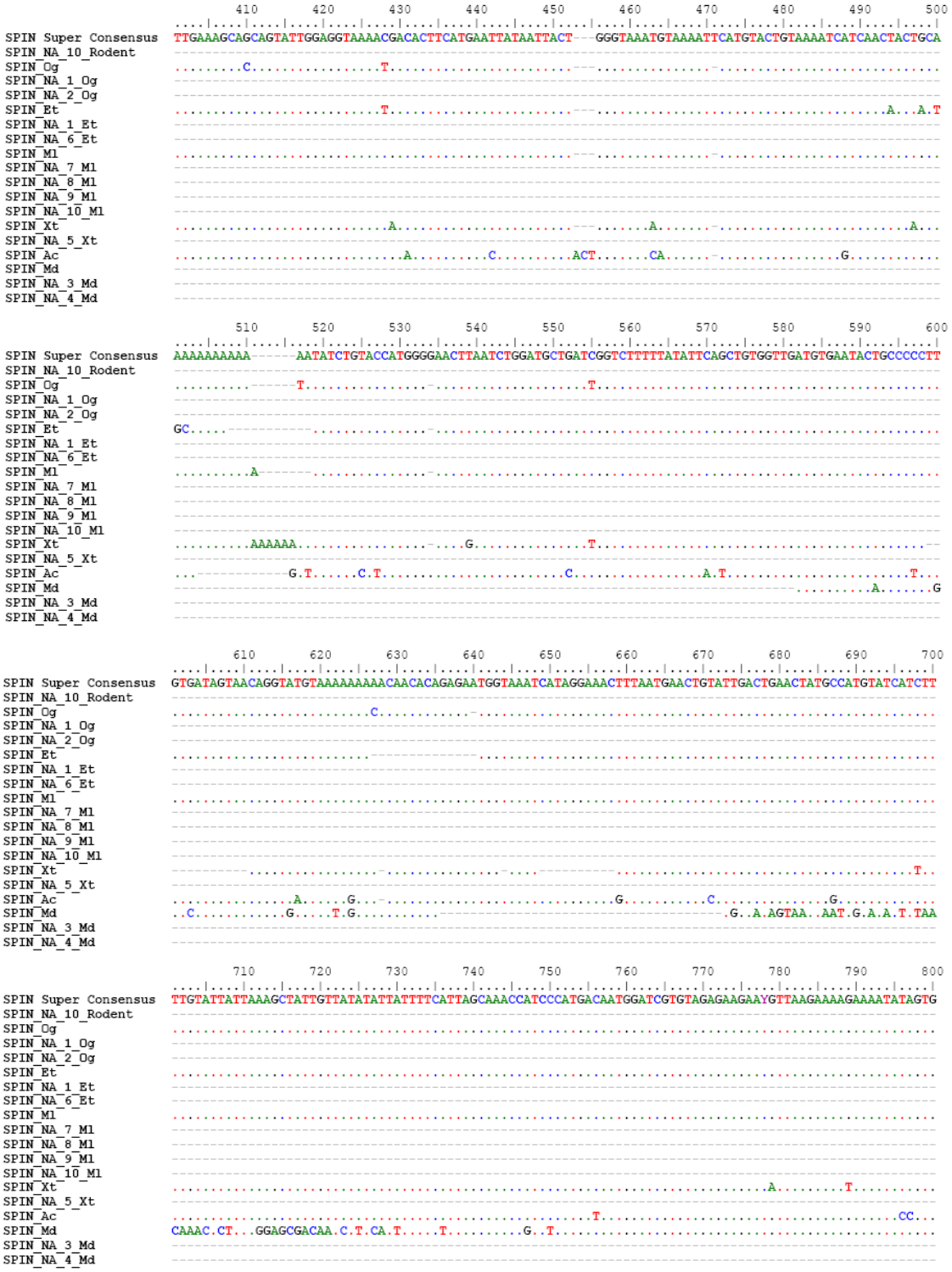
4	MER46C	95	36.9
4	MER5A	96	77.3
4	Kanga1a	96	42.0
4	Charlie7	96	27.4
4	MER91B	97	40.9
4	Kanga1b	97	33.1
4	MER103	97	39.4
4	MER113	97	36.9
4	Kanga2_a	98	41.9
4	L1PB4	98	30.8
4	MER20B	98	42.7
4	FordPrefect_a	98	57.5
4	MER91C	100	25.6
4	L1PA17	100	68.6
4	L1MA5	101	28.2
4	Charlie2b	101	45.7
4	Charlie8	101	40.5
4	L1MA4	101	34.6
4	MER5B	101	53.0
4	Tigger8	106	29.1
4	MER115	106	42.3
4	MER102b	106	65.4
4	MER91A	107	45.8
4	MER102a	107	72.6
4	MER117	111	33.2
4	MARNA	111	46.2

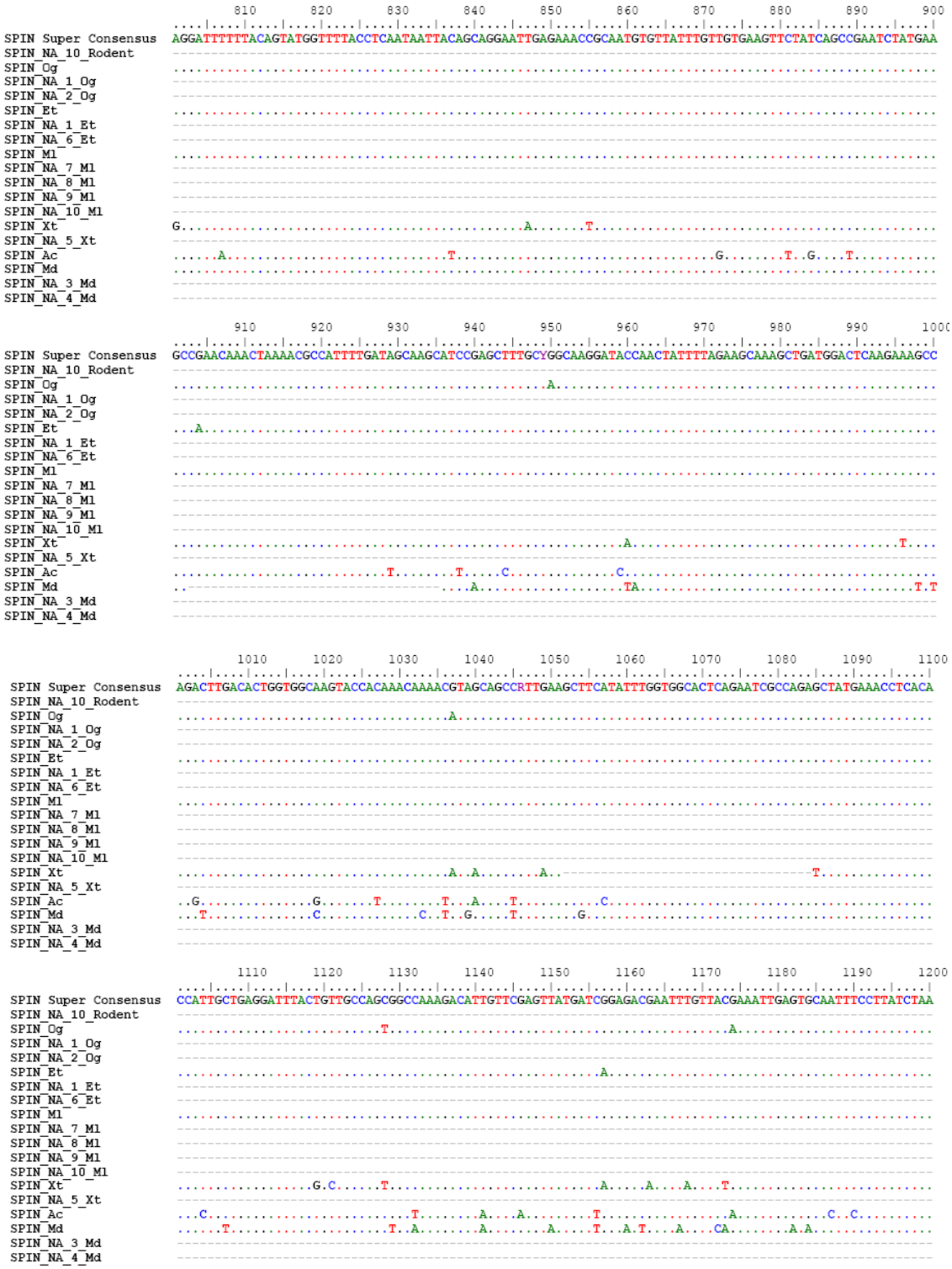
*% Divergence was calculated using PAML 3.15. The REV model was used for the calculations.

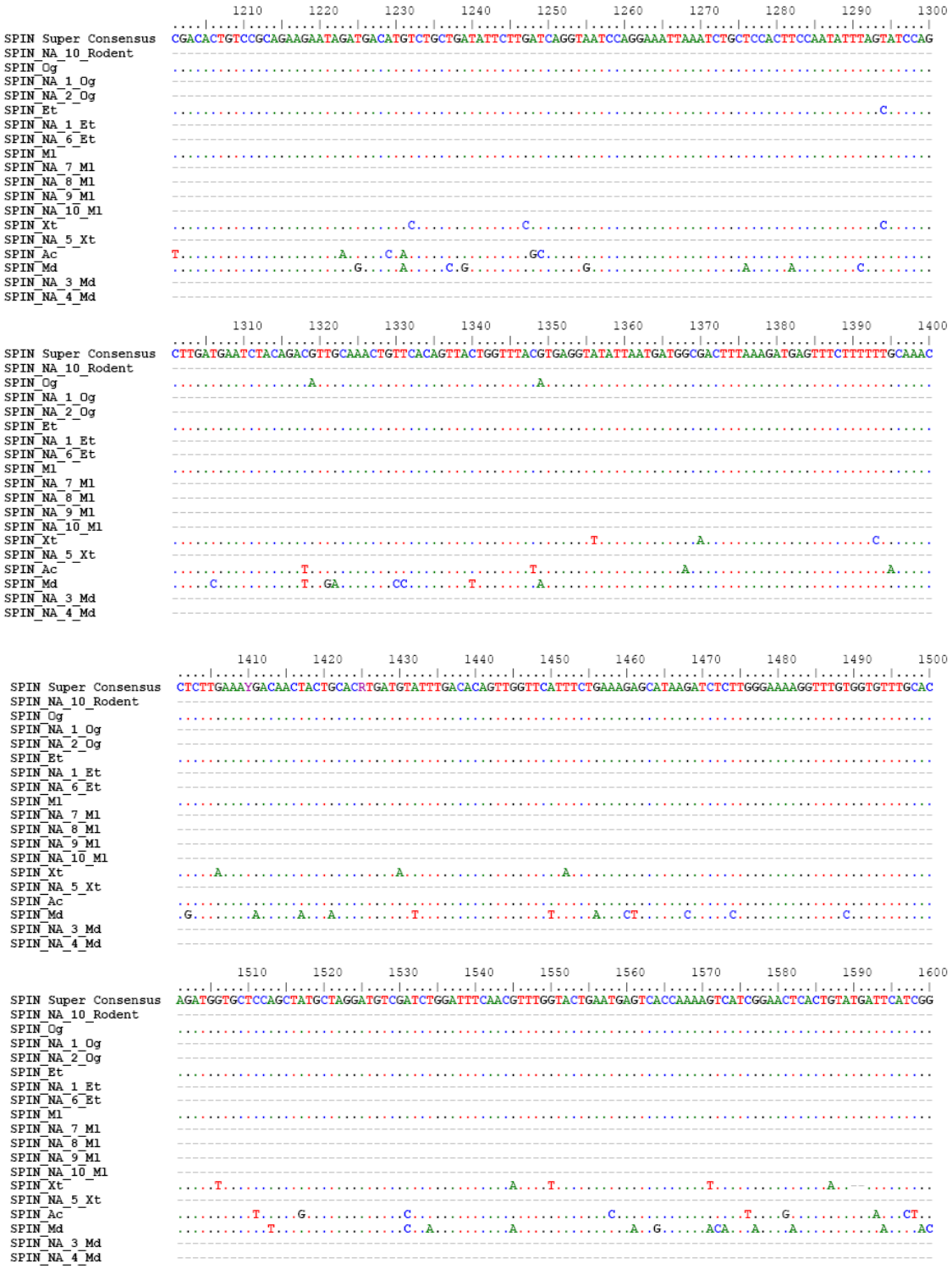
APPENDIX B

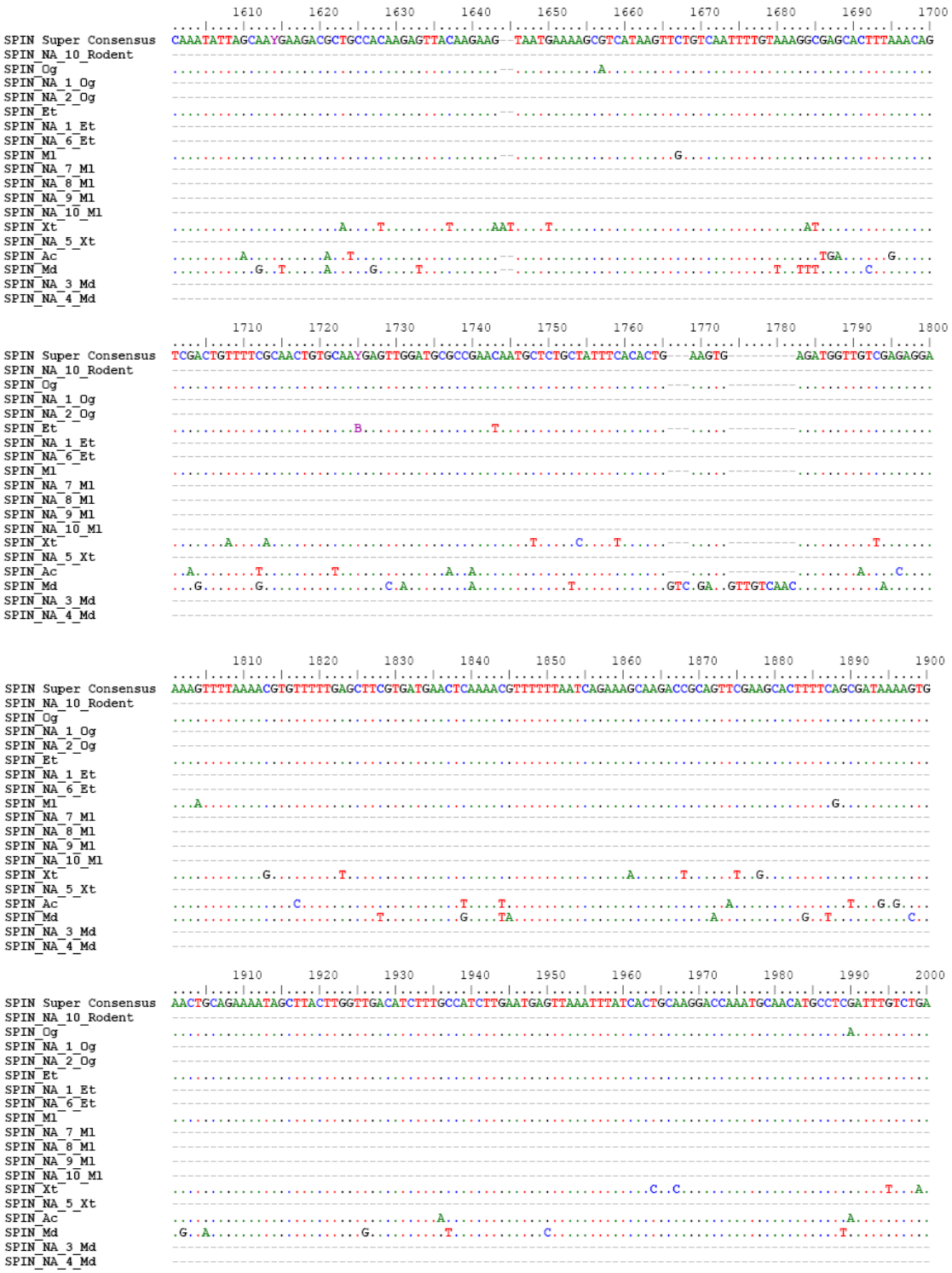
MULTIPLE ALIGNMENT OF *SPIN* TRANSPOSONS AND THEIR DERIVED MITES IN EACH SPECIES. ALL SEQUENCES ARE CONSENSUS SEQUENCES, EXCEPT *SPIN_XT* AND *SPIN_AC*, WHICH ARE INDIVIDUAL *SPIN* COPIES. MITES ARE DENOTED BY THE SUFFIX 'NA' (FOR NONAUTONOMOUS). THE ALIGNMENT SHOWS THAT THE MITE FAMILIES FROM DIFFERENT SPECIES HAVE DISTINCT DELETION BREAKPOINTS WITH THEIR COGNATE FULL-LENGTH *SPIN* TRANSPOSON, INDICATING THAT EACH ORIGINATED INDEPENDENTLY FROM A DISTINCT DELETION DERIVATIVE. SEQUENCES USED ARE INCLUDED IN FASTA FORMAT FOLLOWING THE ALIGNMENT. SPECIES ABBREVIATIONS: RODENT = MOUSE/RAT, OG = OTOLEMUR GARNETTII (BUSHBABY), ET = ECHINOPS TELFAIRI (TENREC), ML = MYOTIS LUCIFUGUS (BAT), XT = XENOPUS TROPICALIS (FROG), AC = ANOLIS CAROLINENSIS (LIZARD), MD = MONODELPHIS DOMESTICA (OPOSSUM).

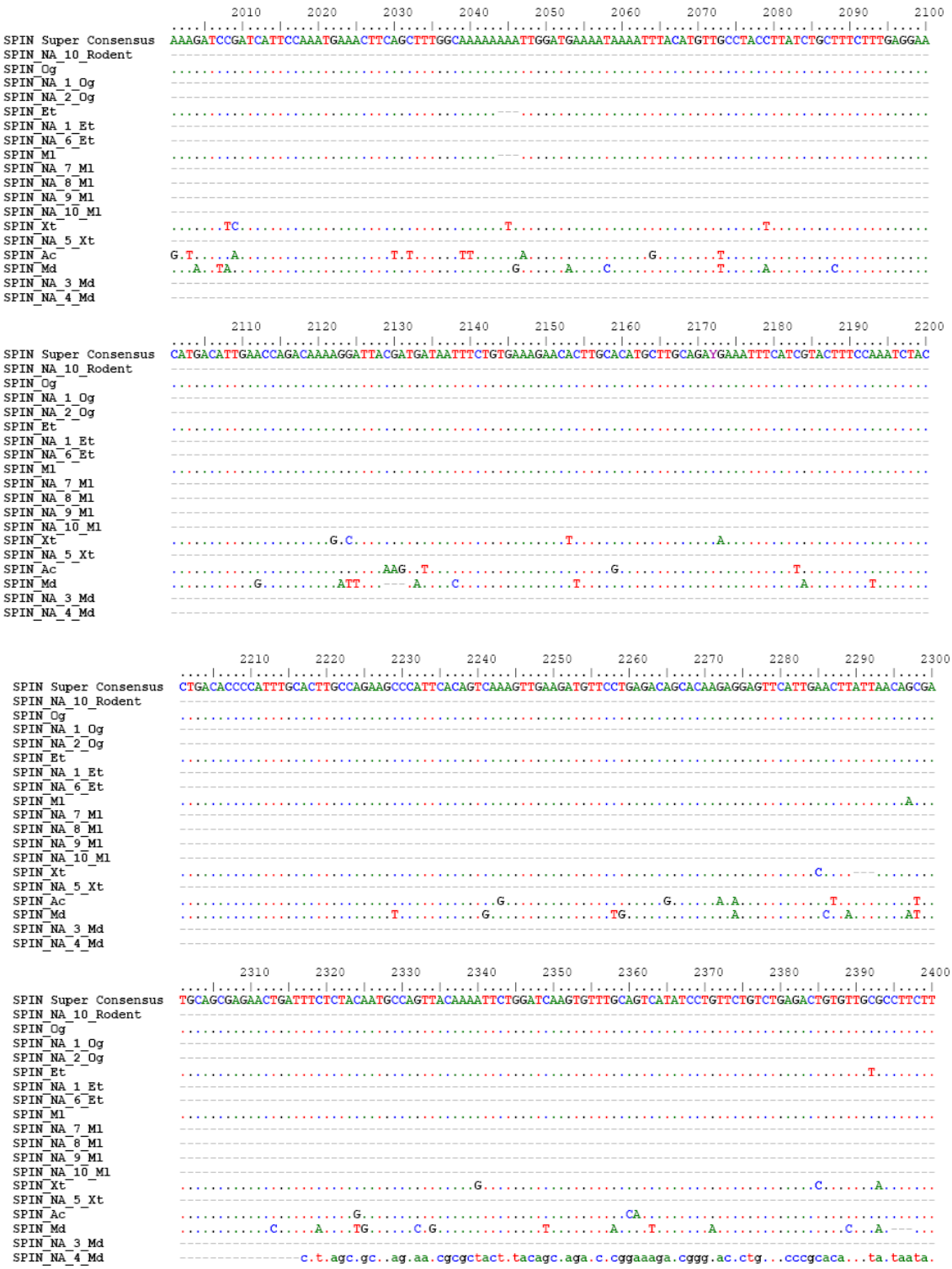

SPIN Super Consensus
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md

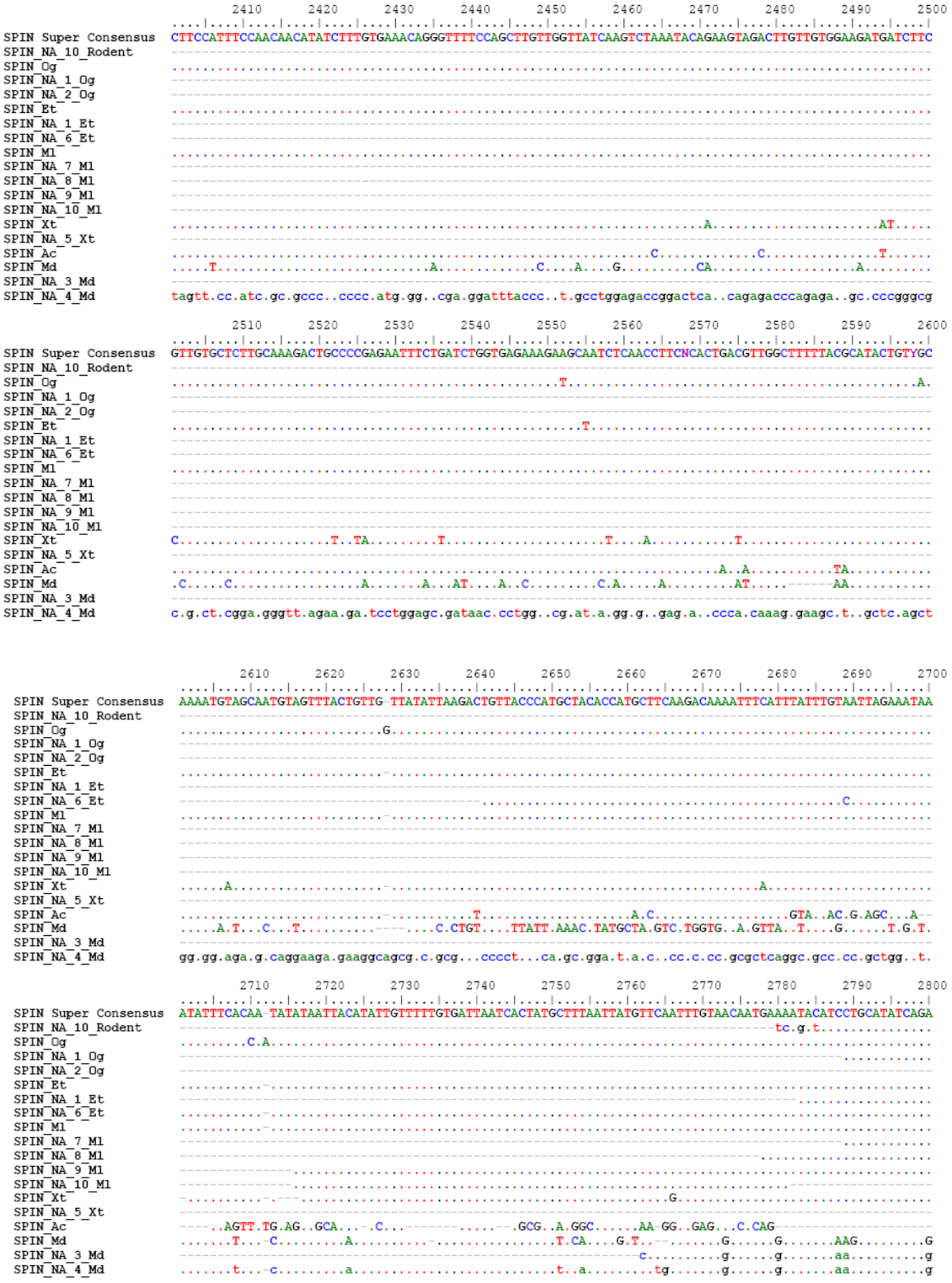

410 420 430 440 450 460 470 480 490 500
SPIN Super Consensus TTGAAAGCAGCAGTATTGGAGGTAAAACGACACTTCATGAATTATAATTACT---GGGTAAATGTAAAATTCATGTACTGTAAAATCATCAACTACTGCA
SPIN_NA_10_Rodent
SPIN_Og .....C.....T.....---.....-.....
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et .....T.....---.....A...A.T
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml .....---.....-.....
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt .....A.....---.....A.....-.....A...
SPIN_NA_5_Xt
SPIN_Ac .....A.....C.....ACT.....CA.....-.....G.....
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
510 520 530 540 550 560 570 580 590 600
SPIN Super Consensus AAAAAAAAAA-----AATATCTGTACCATGGGGAACTTAATCTGGATGCTGATCGGTCTTTTTATATTCAGCTGTGGTTGATGTGAATACTGCCCCCTT
SPIN_NA_10_Rodent
SPIN_Og .....-----T.....-.....T.....
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et GC.....----------.....-.....
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml .....A------.....-.....
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt .....AAAAAA.....-....G.....T.....--
SPIN_NA_5_Xt
SPIN_Ac ...----------G.T.....C.T.....C.....A.T.....T...
SPIN_Md .....A.....G
SPIN_NA_3_Md
SPIN_NA_4_Md
610 620 630 640 650 660 670 680 690 700
SPIN Super Consensus GTGATAGTAACAGGTATGTAAAAAAAAACAACACAGAGAATGGTAAATCATAGGAAACTTTAATGAACTGTATTGACTGAACTATGCCATGTATCATCTT
SPIN_NA_10_Rodent
SPIN_Og .....C.....-.....
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et .....----------.....
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml .....
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt ---------.....-.....-....-.....T..
SPIN_NA_5_Xt
SPIN_Ac .....A.....G...-.....G.....C.....G.....
SPIN_Md ..C.....G.....T.G.....-.....G..A.AGTAA..AAT.G.A.A.T.TAA
SPIN_NA_3_Md
SPIN_NA_4_Md
710 720 730 740 750 760 770 780 790 800
SPIN Super Consensus TTGTATTATTAAAGCTATTGTTATATATTATTTTCATTAGCAAACCATCCCATGACAATGGATCGTGTAGAGAAGAAYGTTAAGAAAAGAAAATATAGTG
SPIN_NA_10_Rodent
SPIN_Og .....
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et .....
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml .....
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt .....A.....T.....
SPIN_NA_5_Xt
SPIN_Ac .....T.....CC...
SPIN_Md CAAAC.CT...GGAGCGACAA.C.T.CA.T.....T.....G..T.....
SPIN_NA_3_Md
SPIN_NA_4_Md

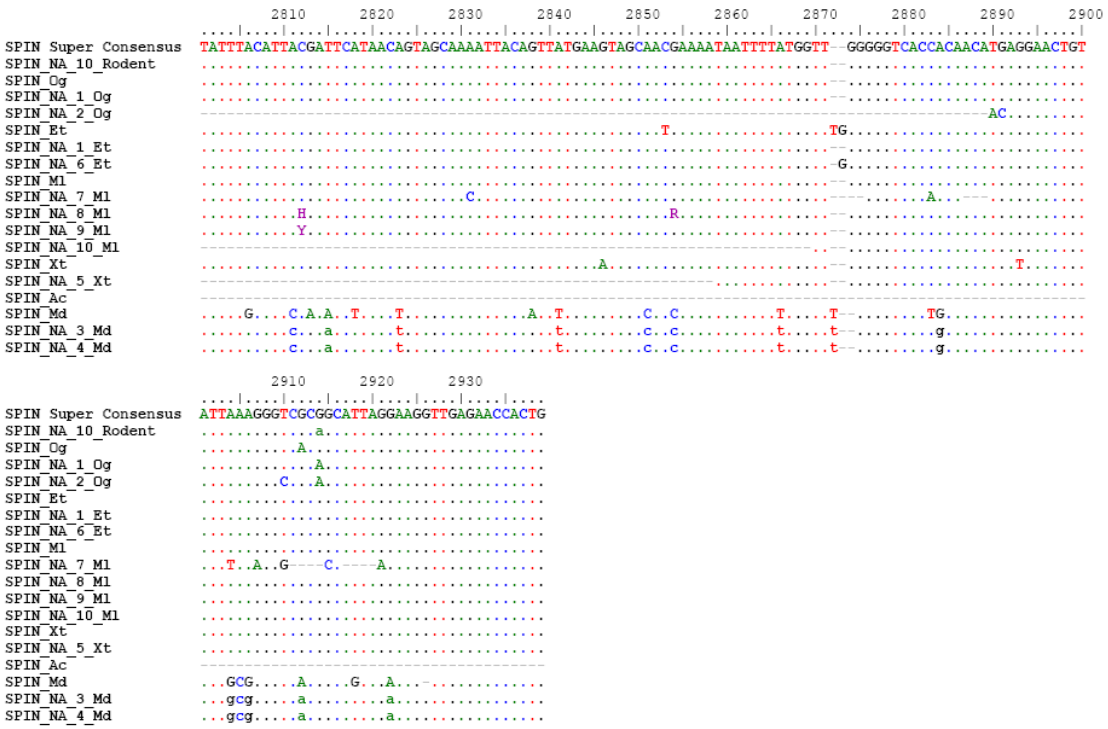

SPIN Super Consensus
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
SPIN Super Consensus CGACACTGTCCGCAGAAGAATAGATGACATGTCTGCTGATATTCTTGATCAGGTAATCCAGGAAATTAAATCTGCTCCACTTCCAATATTTAGTATCCAG
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
SPIN Super Consensus CTTGATGAATCTACAGACGTTGCAAACTGTTCACAGTTACTGGTTTACGTGAGGTATATTAATGATGGCGACTTTAAAGATGAGTTTCTTTTTTGCAAAC
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
SPIN Super Consensus CTCTTGAAAYGACAACTACTGCACRTGATGTATTTGACACAGTTGGTTCATTTCTGAAAGAGCATAAGATCTCTTGGGAAAAGGTTTGTGGTGTTTGCAC
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
SPIN Super Consensus AGATGGTGCTCCAGCTATGCTAGGATGTCGATCTGGATTTCAACGTTTGGTACTGAATGAGTCACCAAAAGTCATCGGAACTCACTGTATGATTCATCGG
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md


1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
SPIN Super Consensus CAAATATTAGCAAYGAAGACGCTGCCACAAGAGTTACAAGAAG--TAATGAAAAGCGTCATAAGTTCTGTCAATTTTGTAAAGGCGAGCACTTTAAACAG
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
SPIN Super Consensus TCGACTGTTTTCGCAACTGTGCAAYGAGTTGGATGCGCCGAACAATGCTCTGCTATTTCACACTG--AAGTG-------AGATGGTTGTCGAGAGGA
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
SPIN Super Consensus AAAGTTTTAAAACGTGTTTTTGAGCTTCGTGATGAACTCAAAACGTTTTTTAATCAGAAAGCAAGACCGCAGTTCGAAGCACTTTTCAGCGATAAAAGTG
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
SPIN Super Consensus AACTGCAGAAAATAGCTTACTTGGTTGACATCTTTGCCATCTTGAATGAGTTAAATTTATCACTGCAAGGACCAAATGCAACATGCCTCGATTTGTCTGA
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md


2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
SPIN Super Consensus AAAGATCCGATCATTCCAAATGAAACTTCAGCTTTGGCAAAAAAAATTGGATGAAAATAAAATTTACATGTTGCCTACCTTATCTGCTTTCTTTGAGGAA
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
SPIN Super Consensus CATGACATTGAACCAGACAAAAGGATTACGATGATAATTTCTGTGAAAGAACACTTGCACATGCTTGCAGAYGAAATTTCATCGTACTTTCCAAATCTAC
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
SPIN Super Consensus CTGACACCCCATTTGCACTTGCCAGAAGCCCATTCACAGTCAAAGTTGAAGATGTTCCTGAGACAGCACAAGAGGAGTTCATTGAACTTATTAACAGCGA
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
SPIN Super Consensus TGCAGCGAGAACTGATTTCTCTACAATGCCAGTTACAAAATTCTGGATCAAGTGTTTGCAGTCATATCCTGTTCTGTCTGAGACTGTGTTGCGCCTTCTT
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md


2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
SPIN Super Consensus CTTCCATTTCCAACAACATATCTTTGTGAAACAGGGTTTTCCAGCTTGTTGGTTATCAAGTCTAAATACAGAAGTAGACTTGTTGTGGAAGATGATCTTC
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md tagtt.cc.atc.gc.gccc..cccc.atg.gg..cga.ggatttaccc..t.gcctggagaccggactca..cagagacccagaga..gc.cccgggcg
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
SPIN Super Consensus GTTGTGCTCTTGCAAAGACTGCCCCGAGAATTTCTGATCTGGTGAGAAAGAAGCAATCTCAACCTTCNCACTGACGTTGGCTTTTTACGCATACTGTYGC
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md c.g.ct.cgga.gggtt.agaa.ga.tcctggagc.gataac.cctgg..cg.at.a.gg.g..gag.a..ccca.caaag.gaagc.t..gctc.agct
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
SPIN Super Consensus AAAATGTAGCAATGTAGTTTACTGTTG TTATATTAAGACTGTTACCCATGCTACACCATGCTTCAAGACAAAATTTCATTTATTTGTAATTAGAAATAA
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md gg.gg.aga.g.caggaaga.gaaggcagcg.c.gcg...cccct...ca.gc.gga.t.a.c..cc.c.cc.gcgctcaggc.gcc.cc.gctgg..t.
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
SPIN Super Consensus ATATTTCACAA TATATAATTACATATTGTTTTTGTGATTAATCACTATGCTTTAATTATGTTCAATTTGTAACAATGAAAATACATCCTGCATATCAGA
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md
SPIN_NA_3_Md
SPIN_NA_4_Md


2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
SPIN Super Consensus TATTTACATTACGATTCATAACAGTAGCAAAATTACAGTTATGAAGTAGCAACGAAAATAATTTTATGGTT GGGGGTCACCACAACATGAGGAACTGT
SPIN_NA_10_Rodent
SPIN_Og
SPIN_NA_1_Og
SPIN_NA_2_Og AC
SPIN_Et T TG
SPIN_NA_1_Et
SPIN_NA_6_Et G
SPIN_Ml
SPIN_NA_7_Ml C A
SPIN_NA_8_Ml H R
SPIN_NA_9_Ml Y
SPIN_NA_10_Ml
SPIN_Xt A T
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md G C A A T T A T C C T T TG
SPIN_NA_3_Md c a t t c c t t g
SPIN_NA_4_Md c a t t c c t t g
2910 2920 2930
SPIN Super Consensus ATTAAAGGGTCGCGGCATTAGGAAGGTTGAGAACCACTG
SPIN_NA_10_Rodent a
SPIN_Og A
SPIN_NA_1_Og A
SPIN_NA_2_Og C A
SPIN_Et
SPIN_NA_1_Et
SPIN_NA_6_Et
SPIN_Ml
SPIN_NA_7_Ml T A G C A
SPIN_NA_8_Ml
SPIN_NA_9_Ml
SPIN_NA_10_Ml
SPIN_Xt
SPIN_NA_5_Xt
SPIN_Ac
SPIN_Md GCG A G A
SPIN_NA_3_Md gcg a a
SPIN_NA_4_Md gcg a a

APPENDIX C

DNA TRANSPOSON PHYLOGENY
EACH TRIBE OF TES IS GROUPED BY THE AUTONOMOUS ELEMENT (IN BOLD) WITH THE CORRESPONDING MITE FAMILIES (INDENTED BELOW). ORPHAN TE FAMILIES ARE ITALICIZED. THE CORRESPONDING REGIONS COLUMN LISTS THE PORTIONS OF THE ELEMENTS THAT ALIGN WITH EACH OTHER. EACH SEPARATE TRIBE IS DELINEATED BY A BLACK RECTANGLE.

TE Name	Classification	Autonomous/ Nonautonomous	Superfamily	Length	Corresponding regions
Arthur1	Eutherian	A	hAT	3947	
Arthur1A	Eutherian	N	hAT	177	Arthur1: 1-79, 3849-3947 Arthur1A: 1-79, 79-177
Arthur1B	Eutherian	N	hAT	1225	Arthur1: 1-836, 3551-3947 Arthur1B: 1-835, 829-1225
Arthur1C	Eutherian	N	hAT	363	Arthur1: 1-67, 3801-3947 Arthur1C: 1-67, 217-363
Blackjack	Eutherian	A	hAT	2944	
MER81	Eutherian	N	hAT	114	Blackjack: 1-65, 33-1 MER81: 1-67, 82-114
MER94	Eutherian	N	hAT	134	Blackjack: 1-60, 2873-2944 MER94: 1-60, 63-134
Charlie1	Eutherian	A	hAT	2781	
Charlie1a	Eutherian	N	hAT	1450	Charlie1: 1-118, 1442-2781 Charlie1a: 1-118, 116-1450
Charlie1b	Eutherian	N	hAT	518	Charlie1: 1-177, 2430-2781 Charlie1b: 1-177, 172-518
Charlie1b_Mars	Eutherian	N	hAT	1762	Charlie1: 1-172, 1116-2781 Charlie1b_Mars: 1-172, 148-1762
Charlie2	Eutherian	A	hAT	2760	
Charlie2a	Eutherian	A	hAT	2862	Charlie2: 1-607, 1610-2760 Charlie2a: 1-607, 1716-2862
Charlie2b	Eutherian	A	hAT	2782	
Charlie3	Anthropoid	A	hAT	2710	
MER1A	Anthropoid	N	hAT	527	Charlie3:1-274, 2456-2710 MER1A: 1-274, 273-527
MER1B	Anthropoid	N	hAT	337	Charlie3: 1-96, 135-281, 2621-2710 MER1B: 1-96, 95-241, 248-337
Charlie3_Bt	Cow	A	hAT	2713	Charlie3: 1-2710 Charlie3_Bt: 1-2713
MER1A_Bt_1	Cow	N	hAT	193	Charlie3_Bt: 1-35, 155-185, 2578-2713 MER1A_Bt_1: 1-35, 35-65, 58-193

MER1A_Bt_2	Cow	N	hAT	291	Charlie3_Bt: 1-196, 2475-2713 MER1A_Bt_2: 1-197, 197-434
Charlie3_Og	Prosimian	A	hAT	2727	Charlie3: 5-2710 Charlie3_Og: 5-2727
MER1A_Og_1	Prosimian	N	hAT	327	Charlie3_Og: 1-268, 2665-2727 MER1A_Og_1: 1-266, 266-327
MER1A_Og_2	Prosimian	N	hAT	336	Charlie3_Og: 1-133, 2536-2727 MER1A_Og_2: 1-137, 138-336
MER1A_Og_3	Prosimian	N	hAT	541	Charlie3_Og: 1-310, 2497-2727 MER1A_Og_3: 1-310, 307-541
MER1A_Og_4	Prosimian	N	hAT	488	Charlie3_Og: 1-401, 2612-2727 MER1A_Og_4: 1-407, 370-488
MER1A_Og_5	Prosimian	N	hAT	393	Charlie3_Og: 1-288, 2620-2727 MER1A_Og_5: 1-286, 284-393
MER1A_Tb_1	Treeshrew	N	hAT	434	Charlie3: 1-196, 2474-2710 MER1A_Tb_1: 1-197, 197-434
MER1A_Tb_2	Treeshrew	N	hAT	339	Charlie3: 1-245, 2629-2710 MER1A_Tb_2: 1-249, 254-339
MER1A_Tb_3	Treeshrew	N	hAT	335	Charlie3: 1-51, 89-210, 2543-2710 MER1A_Tb_3: 1-51, 52-173, 172-335
MER1A_Tb_4	Treeshrew	N	hAT	513	Charlie3: 1-329, 2526-2710 MER1A_Tb_4: 1-330, 320-513
Charlie4	Eutherian	A	hAT	1961	
Charlie4a	Eutherian	N	hAT	480	Charlie4: 1-291, 1743-1933 Charlie4a: 1-291, 290-480
Charlie4z	Eutherian	N	hAT	167	Charlie4: 1-51, 1883-1961 Charlie4z: 1-51, 88-167
Charlie5	Eutherian	A	hAT	2612	
MER3	Eutherian	N	hAT	209	Charlie5: 1-118, 1-56 MER3: 48-168, 209-156
MER33	Eutherian	N	hAT	324	Charlie5: 1-65, 2343-2612 MER33: 1-65, 53-324
Charlie6	Eutherian	A	hAT	3500	
Charlie7	Eutherian	A	hAT	2612	
Charlie7a	Eutherian	N	hAT	272	Charlie7: 1-112, 2420-2612 Charlie7a: 1-114, 70-272
Charlie8	Eutherian	A	hAT	2417	
MER102a	Eutherian	N	hAT	330	Charlie8: 1-273, 2351-2417 MER102a: 1-273, 264-330
MER102b	Eutherian	N	hAT	341	Charlie8: 1-273, 2351-2417 MER102b: 1-284, 275-341

MER102c	Eutherian	N	hAT	331	Charlie8: 1-68, 92-265, 2341-2417 MER102c: 1-79, 73-246, 255-331
Charlie9	Eutherian	A	hAT	2757	
MER112	Eutherian	N	hAT	261	Charlie9: 1-81, 2630-2757 MER112: 1-81, 123-251
Charlie10	Eutherian	A	hAT	2822	
Charlie10A	Eutherian	N	hAT	280	Charlie10: 1-76, 1374-1485, 1374-1407, 1-76 Charlie10A: 1-76, 70-180, 211-179, 280-205
Charlie10B	Eutherian	N	hAT	242	Charlie10: 1-24, 1374-1485, 1374-1407, 1-76 Charlie10B: 1-25, 31-141, 172-140, 242-166
MER5C	Eutherian	N	hAT	324	Charlie10: 1-124, 2633-2822 MER5C: 1-124, 128-324
MER5C1	Eutherian	N	hAT	263	Charlie10: 1-151, 2706-2822 MER5C1: 1-124, 128-324
Charlie11	Eutherian	A	hAT	2196	
Charlie12	Anthropoid	A	hAT	2873	
MER30	Anthropoid	N	hAT	230	Charlie12: 1-126, 2764-2873 MER30: 1-126, 118-230
MER30b	Anthropoid	N	hAT	190	Charlie12: 2705-2873 MER30b: 22-190
MER30_MI_1	Bat	N	hAT	80	Charlie12: 1-36 MER30_MI_1: 1-36
MER30_MI_2	Bat	N	hAT	246	Charlie12: 1-64 MER30_MI_2: 1-65
MER30_MI_3	Bat	N	hAT	621	Charlie12: 1-42 MER30_MI_3: 1-42
MER30_Og_1	Bat	N	hAT	208	Charlie12: 7-36, 2709-2873 MER30_Og_1: 7-36, 46-208
RMER30	Guinea pig, murine rodents	N	hAT	201	Charlie12: 1-173 RMER30: 1-164
Charlie13a	Eutherian	N	hAT	1508	
Charlie13b	Eutherian	N	hAT	506	Charlie13a: 1-126, 1415-1508 Charlie13b: 1-126, 413-506
Charlie14a	Eutherian	N	hAT	1164	
Charlie15a	Eutherian	N	hAT	224	

Charlie16a	Eutherian	N	hAT	342	
Charlie17a	Eutherian	N	hAT	219	
Charlie18a	Eutherian	N	hAT	342	
Charlie19a	Eutherian	N	hAT		
Charlie20a	Eutherian	N	hAT	807	
Charlie21a	Eutherian	N	hAT	1213	
Charlie22a	Eutherian	N	hAT	491	
Charlie23a	Eutherian	N	hAT	339	
Charlie24	Eutherian	A	hAT	2450	
Charlie25	Eutherian	A	hAT	2524	
Charlie26a	Eutherian	N	hAT	325	
Cheshire	Eutherian	A	hAT	2285	
Cheshire_Mars	Eutherian	A	hAT	2404	
CheshMITE	Eutherian	N	hAT	205	Cheshire: 1-95, 1-95 Chesh_MITE: 1-97, 205-111
MER58A	Eutherian	N	hAT	224	Cheshire: 1-57, 173-280, 2358-2420 MER58A: 1-57, 56-163, 162-224
MER58B	Eutherian	N	hAT	341	Cheshire: 1-280, 2358-2420 MER58B: 1-280, 279-341
MER58C	Eutherian	N	hAT	215	Cheshire: 1-121, 2358-2420 MER58C: 1-122, 153-215
MER58D	Eutherian	N	hAT	386	Cheshire: 1-86, 2115-2420 MER58D: 1-86, 81-386
Dumbo	Afrotherian	A	Mariner	1614	
Dumbo_NA_1	Afrotherian	N	Mariner	297	Dumbo: 1-106, 1417-1614 Dumbo_NA_1: 1-109, 108-297
Dumbo_NA_2	Afrotherian	N	Mariner	736	Dumbo: 1-197, 1086-1614 Dumbo_NA_2: 1-203, 210-736
FordPrefect	Eutherian	N	hAT	1683	
FordPrefect_a	Eutherian	N	hAT	508	FordPrefect: 1-435, 1609-1683 FordPrefect_a: 1-435, 434-508
HSMAR1	Anthropoid	A	Mariner	1287	
MADE1	Anthropoid	N	Mariner	80	HSMAR1: 1-36, 1242-1287 MADE1: 1-37, 45-1
HSMAR1_Tb	Treeshrew	A	Mariner	1273	HSMAR1: 1-1287 HSMAR1_Tb: 1-1273
HSMAR1_Og	Prosimian	A	Mariner	1287	HSMAR1: 1-1287 HSMAR1_Og: 1-1287
HSMAR1_Bt	Cow	A	Mariner	1286	HSMAR1: 1-1287 HSMAR1_Bt: 1-1286

HSMAR1_Dn	Armadillo	A	Mariner	1286	HSMAR1: 1-1287 HSMAR1_Dn: 1-1286
MMAR1	Murine rodents	A	Mariner	1286	HSMAR1: 1-1287 MMAR1: 1-1286
HSMAR2	Anthropoid	A	Mariner	1301	
HSMAR2_Tb	Treeshrew	A	Mariner	1305	HSMAR2: 1-1299 HSMAR2_Tb: 1-1303
HSMAR2_NA_1_Tb	Treeshrew	N	Mariner	131	HSMAR2_Tb: 3-96, 1260-1303 HSMAR2_Tb_1: 3-93, 87-129
HSMAR2_NA_2_Op	Treeshrew	N	Mariner	131	HSMAR2: 3-87, 1257-1301 HSMAR2_Op_1: 3-86, 86-131
HSMAR2_Oc	Rabbit	A	Mariner	1301	HSMAR2: 1-1300 HSMAR2_Oc: 1-1300
HSMAR2_NA_3_Oc	Rabbit	N	Mariner	132	HSMAR2_Oc: 2-90, 1257-1301 HSMAR2_Oc_1: 2-90, 87-132
HSMAR2_Og	Prosimian	A	Mariner	1303	HSMAR2: 3-1301 HSMAR2_Tb: 3-1299
HSMAR2_Et	Tenrec	A	Mariner	1301	HSMAR2: 3-1299 HSMAR2_Et: 3-1299
HSMAR2_NA_4_Et	Tenrec	N	Mariner	81	HSMAR2_Et: 3-39, 1257-1299 HSMAR2_Et_1: 3-39, 36-79
Kanga1	Eutherian	A	Mariner	1745	
Kanga1a	Eutherian	N	Mariner	754	Kanga1: 1-156, 1138-1745 Kanga1a: 1-156, 146-754
Kanga1b	Eutherian	N	Mariner	779	Kanga1: 1-151, 1365-1745 Kanga1b: 1-151, 399-779
Kanga1c	Eutherian	N	Mariner	729	Kanga1: 1-634, 1639-1745 Kanga1c: 1-631, 623-729
Kanga1d	Eutherian	N	Mariner	609	Kanga1: 1-154, 1283-1745 Kanga1d: 1-153, 147-609
MER104	Eutherian	N	Mariner	181	Kanga1: 1-86, 1647-1745 MER104: 1-84, 83-181
Kanga2_a	Eutherian	A	Mariner	888	
Kanga11a	Eutherian	N	Mariner	970	
LAMAR2_La	Elephant	A	Mariner	931	
LAMAR2_Et	Tenrec	A	Mariner	924	LAMAR2_La: 2-930 LAMAR2_Et: 1-924
LAMAR2_NA_1_Et	Tenrec	N	Mariner	778	LAMAR2_Et: 1-165, 352-924 LAMAR2_Et_1: 1-166, 209-778
Looper	Eutherian	A	piggyBac	1558	
MADE2	Eutherian	N	Mariner	80	
MamRep38	Eutherian	N	hAT	295	

MamRep1894	Eutherian	N	hAT	123	MamRep38: 1-74, 75-123 MamRep1894: 1-74, 247-295
MamRep137	Eutherian	N	Mariner	444	
MamRep434	Eutherian	N	Mariner	426	
MamRep1161	Eutherian	N	Mariner	957	
Tigger16A	Eutherian	N	Mariner	933	MamRep1161: 1-957 Tigger16A: 6-928
Tigger16B	Eutherian	N	Mariner	337	MamRep1161: 787-957 Tigger16B: 162-332
MamRep1879	Eutherian	N	hAT	216	
MamRep4096	Eutherian	N	hAT	426	
MARNA	Eutherian	N	Mariner	586	
MER2	Eutherian	N	Mariner	344	
MER2B	Eutherian	N	Mariner	366	MER2: 1-77, 172-344 MER2B: 1-77, 164-336
MER5A	Eutherian	N	hAT	189	
MER5A1	Eutherian	N	hAT	166	MER5A: 4-189 MER5A1: 4-189
MER5B	Eutherian	N	hAT	178	MER5A: 32-189 MER5B: 22-178
MER6	Eutherian	N	Mariner	865	
MER6A	Eutherian	N	Mariner	605	MER6: 1-347, 597-865 MER6A: 1-347, 337-605
MER6A_MI_1	Eutherian	N	Mariner	608	MER6A: 1-605 MER6A_MI_1: 1-608
MER6B	Eutherian	N	Mariner	210	MER6: 800-865, 1-74, 784-865 MER6B: 66-1, 67-140, 129-210
MER6C	Eutherian	N	Mariner	202	MER6: 800-865, 738-865 MER6C: 66-1, 70-202
MER6_MI_1	Eutherian	N	Mariner	179	MER6: 1-60, 739-865 MER6_MI_1: 1-60, 60-179
MER6_Og_1	Eutherian	N	Mariner	146	MER6: 1-65, 784-865 MER6_Og_1: 1-66, 63-146
MER6_Ee_1	Eutherian	N	Mariner	186	MER6: 1-81, 789-865 MER6_Ee_1: 1-78, 110-186
MER20B	Eutherian	N	hAT	783	
MER20	Eutherian	N	hAT	219	MER20B: 1-179 MER20: 1-183
MER45A	Eutherian	N	hAT	178	
MER45B	Eutherian	N	hAT	1040	MER45A: 1-116, 113-178 MER45B: 1-116, 975-1040

MER45C	Eutherian	N	hAT	953	MER45A: 1-116, 113-178 MER45C: 1-118, 888-953
MER45R	Eutherian	N	hAT	1581	MER45A: 1-116, 113-178 MER45D: 1-116, 1516-1581
MER46C	Eutherian	N	Mariner	338	
MER53	Eutherian	N	hAT	193	
MER63A	Eutherian	N	hAT	210	
MER63B	Eutherian	N	hAT	436	MER63A: 1-48, 53-210 MER63B: 1-48, 279-436
MER63C	Eutherian	N	hAT	930	MER63A: 1-48, 53-167, 162-210 MER63C: 1-48, 279-393, 882-930
MER63D	Eutherian	N	hAT	1061	MER63A: 1-48, 53-167, 162-210 MER63D: 1-48, 279-393, 1013-1061
MER75	Anthropoid	N	piggyBac	514	
MER75B	Anthropoid	N	piggyBac	243	MER75: 1-29, 359-514 MER75B: 1-29, 88-243
MER82	Eutherian	N	Mariner	653	
MER85	Anthropoid	N	piggyBac	140	
MER85_Og_1	Prosimian	N	piggyBac	234	MER85: 1-43, 39-234 MER85_Og_1: 1-44, 133-234
MER96	Eutherian	N	hAT	175	
MER96b	Eutherian	N	hAT	417	MER96: 1-58, 1-58 MER96B: 1-62, 417-356
MER97a	Eutherian	N	hAT	894	
MER97b	Eutherian	N	hAT	1053	MER97a: 1-784, 781-894 MER97b: 1-784, 940-1053
MER97c	Eutherian	N	hAT	1090	MER97a: 1-755, 781-894 MER97c: 1-755, 977-1090
MER99	Eutherian	N	hAT	831	
MER103	Eutherian	N	hAT	186	
MER105	Eutherian	N	hAT	204	
MER106A	Eutherian	N	hAT	236	
MER106B	Eutherian	N	hAT	1059	MER106A: 1-103, 109-236 MER106B: 1-103, 932-1059
MER107	Anthropoid	N	hAT	200	
MER113	Eutherian	N	hAT	521	
MER113A	Eutherian	N	hAT	314	MER113: 1-68, 389-521 MER113A: 1-68, 189-314

MER117	Eutherian	N	hAT	197	
MER119	Eutherian	N	hAT	583	
Merlin1_Hs	Supraprimate	A	Merlin	1175	
Oamar1	Cow	A	Mariner	1293	
Oamar1_NA_1	Cow	N	Mariner	80	Oamar1: 1-43, 1254-1293 Oamar1_NA_1: 1-41, 41-80
OGMAR1	Prosimian	A	Mariner	1283	
OGMAR2	Prosimian	A	Mariner	925	
OGMAR2_NA_1	Prosimian	N	Mariner	185	OGMAR2: 925-829, 848-925 OGMAR2_NA_1: 3-99, 106-183
OGMAR2_NA_2	Prosimian	N	Mariner	375	OGMAR2: 1-180, 739-925 OGMAR2_NA_2: 5-184, 187-373
OposCharlie1	Opossum	A	hAT	3328	
OposCharlie1_NA_1_Et	Tenrec	N	hAT	225	OposCharlie1: 1-63, 3172-3328 OposCharlie1_NA_1_Et: 1-63, 70-225
OposCharlie1_NA_2_Og	Prosimian	N	hAT	192	OposCharlie1: 1-63, 3261-3328 OposCharlie1_NA_2_Et: 1-63, 125-192
OposCharlie1_NA_3_Ml	Bat	N	hAT	742	OposCharlie1: 1-63, 2969-3328 OposCharlie1_NA_3_Og: 1-63, 383-742
OposCharlie1_NA_4_Ml	Bat	N	hAT	193	OposCharlie1: 1-60, 3278-3328 OposCharlie1_NA_4_Ml: 1-60, 143-193
OposCharlie1_NA_5_Ml	Bat	N	hAT	193	OposCharlie1: 1-63, 3241-3328 OposCharlie1_NA_5_Ml: 1-63, 107-193
OposCharlie1_NA_6_Ml	Bat	N	hAT	524	OposCharlie1: 1-63, 3185-3328 OposCharlie1_NA_6_Ml: 1-63, 380-524
ORSL	Eutherian	N	hAT	275	
ORSL-2a	Eutherian	N	hAT	341	ORSL: 1-79, 197-275 ORSL-2a: 1-79, 259-341
ORSL-2b	Eutherian	N	hAT	508	ORSL: 1-79, 197-275 ORSL-2b: 1-79, 426-508
RCHARR1	Murine	N	hAT	974	

	rodents				
RCHARR1_Et	Tenrec	N	hAT	229	RCHARR1: 1-176, 924-974 RCHARR1_Et: 1-178, 178-229
Ricksha	Eutherian	A	MuDR	3246	
Ricksha_0	Eutherian	N	MuDR	1733	Ricksha: 1-126, 1691-3236 Ricksha_0: 1-125, 178-1723
Ricksha_a	Eutherian	N	MuDR	1181	Ricksha: 1-141, 2202-3246 Ricksha_a: 1-142, 137-1181
Ricksha_b	Eutherian	N	MuDR	1255	Ricksha: 1-613, 2603-3246 Ricksha_b: 1-613, 612-1255
Ricksha_c	Eutherian	N	MuDR	2048	Ricksha: 1-1298, 2495-3246 Ricksha_c: 1-1298, 1297-2048
SPIN_Et	Tenrec	A	hAT	2871	SPIN_Super: 1-2873 SPIN_Et: 1-2871
SPIN_NA_1_Et	Tenrec	N	hAT	224	SPIN_Super: 1-72, 2680-2836 SPIN_NA_1_Et: 1-72, 68-224
SPIN_NA_6_Et	Tenrec	N	hAT	487	SPIN_Super: 1-190, 2570-2873 SPIN_NA_6_Et: 1-190, 183-487
SPIN_Md	Opossum	A	hAT	2361	SPIN_Super: 1-98, 533-621, 679-854, 884-2873 SPIN_Md: 1-98, 98-186, 208-382, 379-2361
SPIN_NA_3_Md	Opossum	N	hAT	192	SPIN_Super: 2699-2873 SPIN_NA_3_Md: 18-192
SPIN_NA_4_Md	Opossum	N	hAT	718	SPIN_Super: 1-119, 2634-2873 SPIN_NA_4_Md: 1-119, 479-718
SPIN_Ml	Bat	A	hAT	2867	SPIN_Super: 1-2873 SPIN_Ml: 1-2867
SPIN_NA_7_Ml	Bat	N	hAT	212	SPIN_Super: 1-91, 2721-2873 SPIN_NA_7_Ml: 1-93, 73-212
SPIN_NA_8_Ml	Bat	N	hAT	192	SPIN_Super: 1-32, 2708-2873 SPIN_NA_8_Ml: 1-32, 27-192
SPIN_NA_9_Ml	Bat	N	hAT	311	SPIN_Super: 1-89, 2644-2873 SPIN_NA_9_Ml: 1-89, 83-311
SPIN_NA_10_Ml	Bat	N	hAT	223	SPIN_Super: 1-89, 2644-2717, 2806-2873 SPIN_NA_10_Ml: 1-89, 83-155, 156-223

SPIN_NA_10_Rode	Murine rodents	N	hAT	226	SPIN_Super: 1-68, 2722-2873 SPIN_NA_10_Rodent: 1-68, 75-226
SPIN_Og	Prosimian	A	hAT	2836	SPIN_Super: 1-2873 SPIN_Og: 1-2836
SPIN_NA_1_Og	Prosimian	N	hAT	225	SPIN_Super: 1-91, 2684-2836 SPIN_NA_1_Og: 1-93, 73-225
SPIN_NA_2_Og	Prosimian	N	hAT	80	SPIN_Super: 1-30, 2807-2873 SPIN_NA_2_Og: 1-30, 16-80
Tigger1	Eutherian	A	Mariner	2418	
Tigger1a_Art	Eutherian	N	Mariner	497	Tigger1: 1-247, 2165-2418 Tigger1a_Art: 1-247, 244-497
Tigger1a_Car	Eutherian	N	Mariner	471	Tigger1: 1-227, 2172-2418 Tigger1a_Car: 1-227, 225-471
Tigger1a_Mars	Eutherian	N	Mariner	834	Tigger1: 1-458, 2028-2418 Tigger1a_Mars: 1-457, 443-834
Tigger2	Eutherian	A	Mariner	2718	
Tigger2f	Eutherian	A	Mariner	3455	Tigger2: 1-2467, 2461-2718 Tigger2f: 1-2468, 3198-3455
MER8	Eutherian	N	Mariner	239	Tigger2: 1-65 MER8: 1-65
Tigger2a	Eutherian	N	Mariner	341	Tigger2: 1-71, 2461-2718 Tigger2a: 1-71, 84-341
Tigger2a_Art	Eutherian	N	Mariner	220	Tigger2: 1-71, 2646-2718 Tigger2a_Art: 1-74, 149-220
Tigger2a_Car	Eutherian	N	Mariner	341	
Tigger2b	Eutherian	N	Mariner	726	Tigger2: 1-578, 2535-2718 Tigger2b: 1-558, 544-726
Tigger2b_Pri	Eutherian	N	Mariner	1068	Tigger2: 1-757, 2461-2718 Tigger2b_Pri: 1-756, 811-1068
Tigger3(Golem)	Eutherian	A	Mariner	3029	
Tigger3b	Eutherian	N	Mariner	1231	Tigger3(Golem): 1-354, 2150-3029 Tigger3b: 1-356, 352-1231
Tigger4(Zombi)	Supraprimate	A	Mariner	2806	
Tigger5	Eutherian	A	Mariner	2406	
MER47B	Eutherian	N	Mariner	418	Tigger5: 1-243, 2224-2406 MER47B: 1-243, 233-418
MER47C	Eutherian	N	Mariner	97	Tigger5: 1-55, 2334-2406 MER47C: 1-55, 25-97
Tigger5a	Eutherian	N	Mariner	366	Tigger5: 1-94, 2133-2406 Tigger5a: 1-94, 93-366

Tigger6a	Eutherian	N	Mariner	1083	
Tigger6b	Eutherian	N	Mariner	1597	Tigger6a: 1-436, 495-1083 Tigger6b: 1-436, 1009-1597
Tigger7	Eutherian	A	Mariner	2487	
MER44A	Eutherian	N	Mariner	339	Tigger7: 1-74, 2218-2487 MER44A: 1-74, 68-339
MER44B	Eutherian	N	Mariner	550	Tigger7: 1-356, 2289-2487 MER44B: 1-356, 352-550
MER44C	Eutherian	N	Mariner	733	Tigger7: 1-660, 2406-2487 MER44C: 1-659, 652-733
MER44D	Eutherian	N	Mariner	705	Tigger7: 1-238, 2016-2487 MER44D: 1-238, 234-705
Tigger8	Eutherian	N	Mariner	666	
Tigger9a	Eutherian	N	Mariner	732	
Tigger9b	Eutherian	N	Mariner	624	Tigger9a: 1-115, 545-732 Tigger9b: 1-115, 445-624
Tigger10	Eutherian	A	Mariner	1843	
Tigger11a	Eutherian	N	Mariner	735	
Tigger12	Eutherian	A	Mariner	1959	
Tigger12a	Eutherian	N	Mariner	791	Tigger12: 1-536, 1759-1959 Tigger12a: 1-536, 591-791
Tigger12c	Eutherian	N	Mariner	734	Tigger12: 1-666, 1892-1959 Tigger12c: 1-666, 667-734
Tigger13a	Eutherian	N	Mariner	771	
Tigger14a	Eutherian	N	Mariner	324	
Tigger15a	Eutherian	N	Mariner	715	
Zaphod	Eutherian	A	hAT	4031	
MER115	Eutherian	N	hAT	693	Zaphod: 1-67, 3873-4031 MER115: 1-67, 535-693
Zaphod2	Eutherian	A	hAT	3524	
MER91A	Eutherian	N	hAT	196	Zaphod2: 1-49, 3464-3524 MER91A: 1-49, 136-196
MER91B	Eutherian	N	hAT	190	Zaphod2: 1-84, 145-196, 3461-3524 MER91B: 1-74, 79-130, 127-190
MER91C	Eutherian	N	hAT	140	Zaphod2: 1-80, 1-80 MER91C: 1-70, 140-71

APPENDIX D

AGE OF DNA TRANSPOSONS PER SPECIES. THIS DOCUMENT IS ALSO AVAILABLE AT http://feschottelab.uta.edu/pace/age_of_tes.doc. THE ONLINE VERSION INCLUDES THE AGES OF LINE, SINE AND LTR ELEMENTS FOR SPECIES WHERE THE AGES COULD BE CONFIDENTLY CALCULATED.

Age of DNA transposons in armadillo

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	28.40	1283	2.6563E-09	107
arthur1a	28.35	651	2.6563E-09	107
arthur1b	29.51	1946	2.6563E-09	111
Arthur1C	26.19	302	2.6563E-09	99
BLACKJACK	30.43	1891	2.6563E-09	115
Charlie1	25.30	996	2.6563E-09	95
Charlie10	27.85	901	2.6563E-09	105
Charlie10a	30.07	192	2.6563E-09	113
Charlie10b	29.58	154	2.6563E-09	111
Charlie13a	33.87	366	2.6563E-09	128
Charlie13b	35.45	257	2.6563E-09	133
Charlie14a	30.59	455	2.6563E-09	115
Charlie15a	35.21	1670	2.6563E-09	133
Charlie16a	33.45	944	2.6563E-09	126
Charlie17a	34.16	745	2.6563E-09	129
Charlie18a	30.14	1366	2.6563E-09	113
Charlie19a	33.25	846	2.6563E-09	125
Charlie1a	24.12	3078	2.6563E-09	91
Charlie1b	23.44	1740	2.6563E-09	88
Charlie1b_Mars	23.37	1617	2.6563E-09	88
Charlie2	29.26	861	2.6563E-09	110
Charlie20a	33.81	525	2.6563E-09	127
Charlie21a	30.33	401	2.6563E-09	114
Charlie22a	33.00	595	2.6563E-09	124
Charlie23a	33.39	747	2.6563E-09	126
Charlie24	32.31	1560	2.6563E-09	122
Charlie25	35.13	758	2.6563E-09	132
Charlie26a	30.41	283	2.6563E-09	114
Charlie2a	30.05	1077	2.6563E-09	113
Charlie2b	28.32	1751	2.6563E-09	107
Charlie4	24.67	770	2.6563E-09	93
Charlie4a	22.51	16614	2.6563E-09	85
Charlie4z	31.15	3754	2.6563E-09	117
Charlie5	22.33	1693	2.6563E-09	84
Charlie6	27.23	189	2.6563E-09	103
Charlie7	31.08	2272	2.6563E-09	117
Charlie7a	28.61	854	2.6563E-09	108
Charlie8	33.54	2050	2.6563E-09	126
Charlie9	27.93	766	2.6563E-09	105
Cheshire	24.50	466	2.6563E-09	92
Cheshire_Mars	25.04	1379	2.6563E-09	94
CheshMITE	27.18	319	2.6563E-09	102
FordPrefect	31.23	401	2.6563E-09	118
FordPrefect_a	36.74	211	2.6563E-09	138
HSMAR1_Dn	12.24	2347	2.6563E-09	46
Kanga1	29.54	579	2.6563E-09	111
Kanga11a	33.02	580	2.6563E-09	124
Kanga1a	29.49	423	2.6563E-09	111

Age of DNA transposons in armadillo

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1b	29.57	249	2.6563E-09	111
Kanga1c	29.07	613	2.6563E-09	109
Kanga1d	26.58	394	2.6563E-09	100
Kanga2_a	30.36	885	2.6563E-09	114
Looper	22.26	315	2.6563E-09	84
MADE2	22.93	1550	2.6563E-09	86
MamRep1161	33.32	852	2.6563E-09	125
MamRep137	35.46	1254	2.6563E-09	134
MamRep1879	32.26	1011	2.6563E-09	121
MamRep1894	35.97	302	2.6563E-09	135
MamRep38	36.00	1104	2.6563E-09	136
MamRep4096	31.04	705	2.6563E-09	117
MamRep434	33.45	1128	2.6563E-09	126
MARNA	36.42	1979	2.6563E-09	137
MER102a	33.77	1595	2.6563E-09	127
MER102b	36.00	2311	2.6563E-09	136
MER102c	35.13	2290	2.6563E-09	132
MER103	29.36	4141	2.6563E-09	111
MER104	28.40	1389	2.6563E-09	107
MER105	24.34	474	2.6563E-09	92
MER106A	26.39	525	2.6563E-09	99
MER106B	23.53	585	2.6563E-09	89
MER112	28.33	2367	2.6563E-09	107
MER113	30.03	2051	2.6563E-09	113
MER113A	30.02	1008	2.6563E-09	113
MER115	36.23	1457	2.6563E-09	136
MER117	34.84	2328	2.6563E-09	131
MER119	26.87	730	2.6563E-09	101
MER20	23.67	10405	2.6563E-09	89
MER20B	31.85	2377	2.6563E-09	120
MER3	24.35	8217	2.6563E-09	92
MER33	21.57	5775	2.6563E-09	81
MER45A	26.66	1933	2.6563E-09	100
MER45B	25.17	1061	2.6563E-09	95
MER45C	32.52	721	2.6563E-09	122
MER45R	25.54	577	2.6563E-09	96
MER46C	29.82	1652	2.6563E-09	112
MER53	19.39	4034	2.6563E-09	73
MER58A	27.17	7979	2.6563E-09	102
MER58B	25.34	4524	2.6563E-09	95
MER58C	27.89	1677	2.6563E-09	105
MER58D	21.62	889	2.6563E-09	81
MER5A	27.67	14522	2.6563E-09	104
MER5A1	25.59	10084	2.6563E-09	96
MER5B	29.79	10928	2.6563E-09	112
MER5C	26.51	842	2.6563E-09	100
MER5C1	29.84	396	2.6563E-09	112
MER63A	31.03	1902	2.6563E-09	117

Age of DNA transposons in armadillo

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER63B	25.50	1326	2.6563E-09	96
MER63C	25.78	417	2.6563E-09	97
MER63D	24.73	1781	2.6563E-09	93
MER81	28.39	1878	2.6563E-09	107
MER82	23.85	1929	2.6563E-09	90
MER91A	36.94	993	2.6563E-09	139
MER91B	30.65	682	2.6563E-09	115
MER91C	29.26	499	2.6563E-09	110
MER94	30.23	2484	2.6563E-09	114
MER96	26.49	721	2.6563E-09	100
MER96B	22.17	1249	2.6563E-09	83
MER97a	27.28	101	2.6563E-09	103
MER97c	28.29	903	2.6563E-09	106
MER99	25.46	203	2.6563E-09	96
ORSL	23.30	785	2.6563E-09	88
ORSL-2a	30.38	177	2.6563E-09	114
ORSL-2b	29.10	407	2.6563E-09	110
Ricksha	30.42	443	2.6563E-09	115
Ricksha_c	28.53	19614	2.6563E-09	107
Tigger1	16.96	4934	2.6563E-09	64
Tigger10	36.38	597	2.6563E-09	137
Tigger11a	33.96	250	2.6563E-09	128
Tigger12	36.02	279	2.6563E-09	136
Tigger12A	34.79	295	2.6563E-09	131
Tigger12c	36.89	757	2.6563E-09	139
Tigger13a	32.96	1821	2.6563E-09	124
Tigger14a	32.97	758	2.6563E-09	124
Tigger15a	35.85	2676	2.6563E-09	135
Tigger16a	33.72	440	2.6563E-09	127
Tigger16b	35.52	517	2.6563E-09	134
Tigger1a_Art	18.47	554	2.6563E-09	70
Tigger1a_Car	19.26	359	2.6563E-09	73
Tigger1a_Mars	17.24	851	2.6563E-09	65
Tigger2	40.43	157	2.6563E-09	152
Tigger2f	38.89	169	2.6563E-09	146
Tigger6a	22.45	404	2.6563E-09	85
Tigger6b	22.51	323	2.6563E-09	85
Tigger7	37.52	232	2.6563E-09	141
Tigger8	35.51	474	2.6563E-09	134
Tigger9a	31.09	600	2.6563E-09	117
Tigger9b	25.42	519	2.6563E-09	96
Zaphod	28.35	1242	2.6563E-09	107
Zaphod2	28.11	443	2.6563E-09	106

Age of DNA transposons in bat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	28.50	1007	2.6920E-09	106
Arthur1A	28.66	501	2.6920E-09	106
Arthur1B	30.26	1264	2.6920E-09	112
Arthur1C	26.51	427	2.6920E-09	98
BLACKJACK	29.43	1020	2.6920E-09	109
Charlie1	24.98	662	2.6920E-09	93
Charlie10	28.08	624	2.6920E-09	104
Charlie10a	30.55	166	2.6920E-09	113
Charlie10b	29.82	169	2.6920E-09	111
Charlie13a	33.89	359	2.6920E-09	126
Charlie13b	34.39	188	2.6920E-09	128
Charlie14a	30.22	370	2.6920E-09	112
Charlie15a	34.69	1457	2.6920E-09	129
Charlie16a	33.23	779	2.6920E-09	123
Charlie17a	29.19	1468	2.6920E-09	108
Charlie18a	30.13	1261	2.6920E-09	112
Charlie19a	32.94	775	2.6920E-09	122
Charlie1a	24.27	2520	2.6920E-09	90
Charlie1b	23.47	1455	2.6920E-09	87
Charlie1b_Mars	22.85	1315	2.6920E-09	85
Charlie2	28.67	649	2.6920E-09	107
Charlie20a	33.07	429	2.6920E-09	123
Charlie21a	30.47	318	2.6920E-09	113
Charlie22a	32.33	491	2.6920E-09	120
Charlie23a	33.75	630	2.6920E-09	125
Charlie24	33.30	770	2.6920E-09	124
Charlie25	35.06	528	2.6920E-09	130
Charlie26a	30.32	273	2.6920E-09	113
Charlie2a	29.27	720	2.6920E-09	109
Charlie2b	27.98	1301	2.6920E-09	104
Charlie4a	25.97	1842	2.6920E-09	96
Charlie4z	31.00	2817	2.6920E-09	115
Charlie5	22.55	1415	2.6920E-09	84
Charlie6	26.73	106	2.6920E-09	99
Charlie7	31.32	1751	2.6920E-09	116
Charlie7a	28.41	661	2.6920E-09	106
Charlie8	33.39	1612	2.6920E-09	124
Charlie9	28.73	744	2.6920E-09	107
Cheshire	23.72	257	2.6920E-09	88
Cheshire_Mars	24.44	925	2.6920E-09	91
CheshMITE	27.66	312	2.6920E-09	103
FordPrefect	30.83	290	2.6920E-09	115
FordPrefect_a	37.09	156	2.6920E-09	138
hat3_ML	3.02	184	2.6920E-09	11
HeliBat_N1	18.57	132177	2.6920E-09	69
HeliBatN2	14.78	34203	2.6920E-09	55
HeliBatN3	13.01	33136	2.6920E-09	48
Kanga1	29.43	402	2.6920E-09	109

Age of DNA transposons in bat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga11a	32.83	436	2.6920E-09	122
Kanga1a	29.65	359	2.6920E-09	110
Kanga1b	30.56	232	2.6920E-09	114
Kanga1c	30.22	553	2.6920E-09	112
Kanga1d	27.87	308	2.6920E-09	104
Kanga2_a	30.18	675	2.6920E-09	112
Looper	22.50	278	2.6920E-09	84
MADE2	23.18	1293	2.6920E-09	86
MamRep1161	33.65	769	2.6920E-09	125
MamRep137	35.11	1053	2.6920E-09	130
MamRep1879	31.76	835	2.6920E-09	118
MamRep1894	34.72	272	2.6920E-09	129
MamRep38	35.26	1231	2.6920E-09	131
MamRep4096	31.05	593	2.6920E-09	115
MamRep434	33.20	1092	2.6920E-09	123
MARNA	36.05	1579	2.6920E-09	134
MER102a	33.48	1384	2.6920E-09	124
MER102b	34.84	1776	2.6920E-09	129
MER102c	34.13	1707	2.6920E-09	127
MER103	29.56	3605	2.6920E-09	110
MER104	28.95	1030	2.6920E-09	108
MER105	24.46	408	2.6920E-09	91
MER106A	26.39	297	2.6920E-09	98
MER106B	23.77	473	2.6920E-09	88
MER112	28.41	1876	2.6920E-09	106
MER113	30.01	1850	2.6920E-09	111
MER113A	29.71	834	2.6920E-09	110
MER115	35.75	1203	2.6920E-09	133
MER117	34.05	2236	2.6920E-09	126
MER119	28.11	631	2.6920E-09	104
MER2	20.23	1417	2.6920E-09	75
MER20	23.64	9708	2.6920E-09	88
MER20B	31.99	1961	2.6920E-09	119
MER2B	24.77	1097	2.6920E-09	92
MER3	24.62	6046	2.6920E-09	91
MER30_MI_1	11.32	1575	2.6920E-09	42
MER30_MI_2	7.88	229	2.6920E-09	29
MER30_MI_3	7.42	3577	2.6920E-09	28
MER33	22.48	5206	2.6920E-09	84
MER44A	15.89	544	2.6920E-09	59
MER44B	22.87	673	2.6920E-09	85
MER44D	23.81	563	2.6920E-09	88
MER45A	27.08	1795	2.6920E-09	101
MER45B	25.05	727	2.6920E-09	93
MER45C	31.90	631	2.6920E-09	119
MER45R	25.53	438	2.6920E-09	95
MER46C	30.34	1437	2.6920E-09	113
MER47B	21.40	261	2.6920E-09	79

Age of DNA transposons in bat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER53	20.39	3395	2.6920E-09	76
MER58A	27.63	6700	2.6920E-09	103
MER58B	26.40	4312	2.6920E-09	98
MER58C	27.80	1274	2.6920E-09	103
MER58D	22.55	790	2.6920E-09	84
MER5A	27.31	12744	2.6920E-09	101
MER5A1	25.76	9162	2.6920E-09	96
MER5B	25.15	16633	2.6920E-09	93
MER5C	27.10	701	2.6920E-09	101
MER5C1	29.50	293	2.6920E-09	110
MER6	20.72	371	2.6920E-09	77
MER6_MI_1	19.95	1469	2.6920E-09	74
MER63A	30.43	1545	2.6920E-09	113
MER63B	26.24	1218	2.6920E-09	97
MER63C	29.36	508	2.6920E-09	109
MER63D	25.44	1592	2.6920E-09	94
MER6A_MI_1	18.35	724	2.6920E-09	68
MER81	28.85	1741	2.6920E-09	107
MER82	24.48	1921	2.6920E-09	91
MER91A	36.60	763	2.6920E-09	136
MER91B	30.87	526	2.6920E-09	115
MER91C	29.31	364	2.6920E-09	109
MER94	30.21	2152	2.6920E-09	112
MER96	26.42	592	2.6920E-09	98
MER96B	23.10	1190	2.6920E-09	86
MER97c	28.65	779	2.6920E-09	106
MER99	25.89	142	2.6920E-09	96
Myotis_hAT1	2.68	459	2.6920E-09	10
Myotis_Mar	5.36	1265	2.6920E-09	20
Myotis_piggyBac	1.15	490	2.6920E-09	4
Myotis_Tc1	3.79	1367	2.6920E-09	14
Myotis_Tc2	5.91	431	2.6920E-09	22
nhAT_1124	2.75	838	2.6920E-09	10
nhAT_186	2.63	1571	2.6920E-09	10
nhAT_239a	3.93	916	2.6920E-09	15
nhAT1	2.53	18295	2.6920E-09	9
nhAT2	2.00	6051	2.6920E-09	7
nhAT2_525	3.12	359	2.6920E-09	12
nhAT2_730	2.91	1415	2.6920E-09	11
nhAT3	4.64	25600	2.6920E-09	17
nhAT3_200	1.99	5304	2.6920E-09	7
nhAT6	1.83	7578	2.6920E-09	7
nMar_1265	5.31	878	2.6920E-09	20
nMar_1285	5.44	550	2.6920E-09	20
npiggy_156	1.48	1517	2.6920E-09	6
npiggy_239	1.18	389	2.6920E-09	4
npiggy2_345	3.15	260	2.6920E-09	12
nTc1_452	2.90	1173	2.6920E-09	11

Age of DNA transposons in bat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
nTc1_950	4.27	3977	2.6920E-09	16
nTc2_527	8.12	246	2.6920E-09	30
nTc2_555	7.20	223	2.6920E-09	27
OposCharlie1	17.73	148	2.6920E-09	66
OposCharlie1_NA_3_MI	3.70	288	2.6920E-09	14
OposCharlie1_NA_4_MI	3.53	415	2.6920E-09	13
OposCharlie1_NA_5_MI	6.29	624	2.6920E-09	23
OposCharlie1_NA_6_MI	3.65	363	2.6920E-09	14
ORSL	23.10	600	2.6920E-09	86
ORSL-2a	30.02	169	2.6920E-09	112
ORSL-2b	29.49	387	2.6920E-09	110
piggyBac2_ML	17.78	4999	2.6920E-09	66
Ricksha	35.86	152	2.6920E-09	133
Ricksha_c	33.33	5862	2.6920E-09	124
SPIN_MI	3.87	2735	2.6920E-09	14
SPIN_NA_10_MI	3.30	11901	2.6920E-09	12
SPIN_NA_7_MI	3.24	21124	2.6920E-09	12
SPIN_NA_8_MI	5.55	3619	2.6920E-09	21
SPIN_NA_9_MI	3.16	10578	2.6920E-09	12
Tigger1_ML	16.12	1461	2.6920E-09	60
Tigger10	37.00	481	2.6920E-09	137
Tigger11a	33.00	211	2.6920E-09	123
Tigger12	34.76	211	2.6920E-09	129
Tigger12A	34.54	286	2.6920E-09	128
Tigger12c	36.48	597	2.6920E-09	135
Tigger13a	32.81	1478	2.6920E-09	122
Tigger14a	32.78	683	2.6920E-09	122
Tigger15a	35.55	2308	2.6920E-09	132
Tigger16a	33.19	378	2.6920E-09	123
Tigger16b	34.82	508	2.6920E-09	129
Tigger1a_Art	19.55	704	2.6920E-09	73
Tigger1a_Car	19.15	1377	2.6920E-09	71
Tigger1a_Mars	17.46	2525	2.6920E-09	65
Tigger2	21.10	231	2.6920E-09	78
Tigger2a	20.02	351	2.6920E-09	74
Tigger2a_Art	26.79	657	2.6920E-09	100
Tigger2a_Car	17.23	1997	2.6920E-09	64
Tigger2b	21.05	292	2.6920E-09	78
Tigger2b_Pri	19.66	707	2.6920E-09	73
Tigger2f	17.75	1002	2.6920E-09	66
Tigger3(Golem)	18.28	4429	2.6920E-09	68
Tigger3b	19.94	611	2.6920E-09	74
Tigger5	19.94	1148	2.6920E-09	74
Tigger6a	23.49	301	2.6920E-09	87
Tigger6b	22.95	244	2.6920E-09	85
Tigger7	19.92	5885	2.6920E-09	74
Tigger8	35.66	403	2.6920E-09	132
Tigger9a	31.20	432	2.6920E-09	116

Age of DNA transposons in bat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger9b	26.12	459	2.6920E-09	97
Zaphod	28.57	971	2.6920E-09	106
Zaphod2	29.43	483	2.6920E-09	109

Age of DNA transposons in bushbaby

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	30.97	1284	2.9590E-09	105
Arthur1A	32.21	106	2.9590E-09	109
Arthur1B	33.07	489	2.9590E-09	112
Arthur1C	29.19	276	2.9590E-09	99
BLACKJACK	31.23	2464	2.9590E-09	106
Charlie1	28.42	933	2.9590E-09	96
Charlie10	30.55	835	2.9590E-09	103
Charlie10a	32.78	180	2.9590E-09	111
Charlie10b	31.71	171	2.9590E-09	107
Charlie12	23.00	3603	2.9590E-09	78
Charlie13a	34.82	349	2.9590E-09	118
Charlie13b	34.69	187	2.9590E-09	117
Charlie14a	31.10	301	2.9590E-09	105
Charlie15a	35.94	1036	2.9590E-09	121
Charlie16a	33.80	683	2.9590E-09	114
Charlie17a	35.28	537	2.9590E-09	119
Charlie18a	31.41	1070	2.9590E-09	106
Charlie19a	33.92	686	2.9590E-09	115
Charlie1a	27.16	3070	2.9590E-09	92
Charlie1b	26.33	1690	2.9590E-09	89
Charlie1b_Mars	25.33	1932	2.9590E-09	86
Charlie2	31.09	889	2.9590E-09	105
Charlie20a	34.71	476	2.9590E-09	117
Charlie21a	32.17	333	2.9590E-09	109
Charlie22a	33.69	458	2.9590E-09	114
Charlie23a	33.83	522	2.9590E-09	114
Charlie24	34.53	778	2.9590E-09	117
Charlie25	36.45	654	2.9590E-09	123
Charlie26a	31.02	264	2.9590E-09	105
Charlie2a	31.23	970	2.9590E-09	106
Charlie2b	30.28	1533	2.9590E-09	102
Charlie3_Og	19.39	1892	2.9590E-09	66
Charlie4	29.60	144	2.9590E-09	100
Charlie4a	27.97	2197	2.9590E-09	95
Charlie4z	32.39	2372	2.9590E-09	109
Charlie5	24.97	1974	2.9590E-09	84
Charlie6	33.25	195	2.9590E-09	112
Charlie7	32.85	1984	2.9590E-09	111
Charlie7a	29.61	628	2.9590E-09	100
Charlie8	34.14	1923	2.9590E-09	115
Charlie9	29.20	706	2.9590E-09	99
Cheshire	27.26	425	2.9590E-09	92
Cheshire_Mars	26.90	1358	2.9590E-09	91
CheshMITE	29.34	323	2.9590E-09	99
FordPrefect	35.20	416	2.9590E-09	119
FordPrefect_a	40.05	162	2.9590E-09	135
HSMAR1_Og	17.51	475	2.9590E-09	59
HSMAR2_Og	19.23	1439	2.9590E-09	65

Age of DNA transposons in bushbaby

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1	32.30	640	2.9590E-09	109
Kanga11a	34.75	465	2.9590E-09	117
Kanga1a	33.25	487	2.9590E-09	112
Kanga1b	31.83	258	2.9590E-09	108
Kanga1c	31.55	581	2.9590E-09	107
Kanga1d	29.47	396	2.9590E-09	100
Kanga2_a	32.10	861	2.9590E-09	108
Looper	26.39	349	2.9590E-09	89
MADE2	24.07	1143	2.9590E-09	81
MamRep1161	33.30	1067	2.9590E-09	113
MamRep137	36.30	994	2.9590E-09	123
MamRep1879	34.30	469	2.9590E-09	116
MamRep1894	36.69	180	2.9590E-09	124
MamRep38	38.22	824	2.9590E-09	129
MamRep4096	32.41	591	2.9590E-09	110
MamRep434	33.36	901	2.9590E-09	113
MARNA	36.97	1504	2.9590E-09	125
MER102a	34.71	1529	2.9590E-09	117
MER102b	37.05	1831	2.9590E-09	125
MER102c	36.04	1797	2.9590E-09	122
MER103	30.56	3135	2.9590E-09	103
MER104	30.38	1097	2.9590E-09	103
MER105	26.97	423	2.9590E-09	91
MER106A	29.55	417	2.9590E-09	100
MER106B	26.30	549	2.9590E-09	89
MER112	30.08	1733	2.9590E-09	102
MER113	31.28	1911	2.9590E-09	106
MER113A	30.73	813	2.9590E-09	104
MER115	38.14	1336	2.9590E-09	129
MER117	35.09	1973	2.9590E-09	119
MER119	30.43	753	2.9590E-09	103
MER1A_Og_1	23.08	1467	2.9590E-09	78
MER1A_Og_2	19.86	1407	2.9590E-09	67
MER1A_Og_3	20.34	1280	2.9590E-09	69
MER1A_Og_4	21.45	2557	2.9590E-09	72
MER1A_Og_5	22.23	2388	2.9590E-09	75
MER2	24.00	5577	2.9590E-09	81
MER20	27.29	10842	2.9590E-09	92
MER20B	34.04	2422	2.9590E-09	115
MER2B	28.20	1829	2.9590E-09	95
MER3	26.97	7371	2.9590E-09	91
MER30_Og_1	19.62	8141	2.9590E-09	66
MER33	24.68	5966	2.9590E-09	83
MER44A	22.38	1213	2.9590E-09	76
MER44B	25.91	851	2.9590E-09	88
MER44C	25.30	361	2.9590E-09	86
MER44D	27.23	569	2.9590E-09	92
MER45A	29.86	1979	2.9590E-09	101

Age of DNA transposons in bushbaby

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER45B	28.38	987	2.9590E-09	96
MER45C	35.85	809	2.9590E-09	121
MER45R	28.47	711	2.9590E-09	96
MER46C	31.93	1656	2.9590E-09	108
MER47B	26.43	375	2.9590E-09	89
MER53	22.17	4136	2.9590E-09	75
MER58A	30.53	7485	2.9590E-09	103
MER58B	28.23	4833	2.9590E-09	95
MER58C	29.70	1507	2.9590E-09	100
MER58D	24.39	846	2.9590E-09	82
MER5A	29.29	13652	2.9590E-09	99
MER5A1	28.01	10087	2.9590E-09	95
MER5B	31.54	9320	2.9590E-09	107
MER5C	28.51	779	2.9590E-09	96
MER5C1	31.84	327	2.9590E-09	108
MER6	24.72	488	2.9590E-09	84
MER6_Og_1	23.67	351	2.9590E-09	80
MER63A	33.15	1720	2.9590E-09	112
MER63B	28.66	1461	2.9590E-09	97
MER63C	28.62	400	2.9590E-09	97
MER63D	27.62	1957	2.9590E-09	93
MER6A	23.54	1218	2.9590E-09	80
MER8	31.32	1055	2.9590E-09	106
MER81	31.17	1882	2.9590E-09	105
MER82	27.73	2270	2.9590E-09	94
MER85_Og_1	13.90	186	2.9590E-09	47
MER91A	38.07	768	2.9590E-09	129
MER91B	31.81	540	2.9590E-09	107
MER91C	29.44	410	2.9590E-09	99
MER94	30.80	2456	2.9590E-09	104
MER96	29.94	540	2.9590E-09	101
MER96B	25.69	1360	2.9590E-09	87
MER97a	29.52	111	2.9590E-09	100
MER97c	30.79	933	2.9590E-09	104
MER99	28.90	212	2.9590E-09	98
OGMAR1	17.70	3120	2.9590E-09	60
OGMAR2	17.10	8848	2.9590E-09	58
OGMAR2_NA_1	18.69	6234	2.9590E-09	63
OGMAR2_NA_2	17.94	11853	2.9590E-09	61
OposCharlie1	11.72	5256	2.9590E-09	40
OposCharlie1_NA_2_Og	10.63	15829	2.9590E-09	36
ORSL	24.75	671	2.9590E-09	84
ORSL-2a	31.49	181	2.9590E-09	106
ORSL-2b	29.96	426	2.9590E-09	101
Ricksha	28.05	679	2.9590E-09	95
Ricksha_b	28.64	177	2.9590E-09	97
Ricksha_c	33.22	10836	2.9590E-09	112
SPIN_NA_1_Og	9.14	8474	2.9590E-09	31

Age of DNA transposons in bushbaby

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
SPIN_NA_2_Og	12.92	17436	2.9590E-09	44
SPIN_Og	7.51	7077	2.9590E-09	25
Tigger1	20.91	9072	2.9590E-09	71
Tigger10	37.84	482	2.9590E-09	128
Tigger11a	35.50	208	2.9590E-09	120
Tigger12	35.75	193	2.9590E-09	121
Tigger12A	35.61	231	2.9590E-09	120
Tigger12c	37.19	483	2.9590E-09	126
Tigger13a	34.35	1606	2.9590E-09	116
Tigger14a	33.26	538	2.9590E-09	112
Tigger15a	35.87	1918	2.9590E-09	121
Tigger16a	33.33	740	2.9590E-09	113
Tigger16b	35.13	496	2.9590E-09	119
Tigger1a_Art	22.65	502	2.9590E-09	77
Tigger1a_Car	23.58	315	2.9590E-09	80
Tigger1a_Mars	22.29	801	2.9590E-09	75
Tigger2	21.83	1443	2.9590E-09	74
Tigger2a	22.48	1823	2.9590E-09	76
Tigger2a_Art	24.18	274	2.9590E-09	82
Tigger2a_Car	22.45	497	2.9590E-09	76
Tigger2b	24.28	257	2.9590E-09	82
Tigger2b_Pri	21.92	780	2.9590E-09	74
Tigger2f	21.69	1382	2.9590E-09	73
Tigger3(Golem)	23.22	9648	2.9590E-09	78
Tigger3b	21.88	3086	2.9590E-09	74
Tigger4(Zombi)	21.05	7623	2.9590E-09	71
Tigger5	25.50	616	2.9590E-09	86
Tigger5a	22.97	3182	2.9590E-09	78
Tigger6a	27.61	428	2.9590E-09	93
Tigger6b	25.89	352	2.9590E-09	88
Tigger7	24.61	3799	2.9590E-09	83
Tigger8	37.31	392	2.9590E-09	126
Tigger9a	33.44	531	2.9590E-09	113
Tigger9b	29.20	543	2.9590E-09	99
Zaphod	31.02	1303	2.9590E-09	105
Zaphod2	30.46	381	2.9590E-09	103

Age of DNA transposons in cat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	29.50	722	2.75E-09	107
Arthur1a	29.78	411	2.75E-09	108
Arthur1b	30.46	1223	2.75E-09	111
Arthur1C	27.72	224	2.75E-09	101
BLACKJACK	30.54	875	2.75E-09	111
Charlie1	25.68	506	2.75E-09	94
Charlie10	29.63	494	2.75E-09	108
Charlie10a	31.74	112	2.75E-09	116
Charlie13a	34.51	211	2.75E-09	126
Charlie13b	34.55	130	2.75E-09	126
Charlie14a	30.24	242	2.75E-09	110
Charlie15a	35.62	814	2.75E-09	130
Charlie16a	33.68	505	2.75E-09	123
Charlie17a	35.68	427	2.75E-09	130
Charlie18a	30.51	688	2.75E-09	111
Charlie19a	33.62	437	2.75E-09	122
Charlie1a	25.26	1905	2.75E-09	92
Charlie1b	24.70	941	2.75E-09	90
Charlie1b_Mars	23.52	928	2.75E-09	86
Charlie2	30.53	564	2.75E-09	111
Charlie20a	34.34	272	2.75E-09	125
Charlie21a	31.02	208	2.75E-09	113
Charlie22a	32.23	336	2.75E-09	117
Charlie23a	33.75	411	2.75E-09	123
Charlie24	33.96	561	2.75E-09	124
Charlie25	35.69	474	2.75E-09	130
Charlie26a	31.37	174	2.75E-09	114
Charlie2a	30.69	590	2.75E-09	112
Charlie2b	28.67	923	2.75E-09	104
Charlie4a	26.85	1273	2.75E-09	98
Charlie4z	31.74	1818	2.75E-09	116
Charlie5	23.27	931	2.75E-09	85
Charlie7	32.03	1295	2.75E-09	117
Charlie7a	29.64	428	2.75E-09	108
Charlie8	34.23	1001	2.75E-09	125
Charlie9	29.00	450	2.75E-09	106
Cheshire	25.48	239	2.75E-09	93
Cheshire_Mars	25.80	663	2.75E-09	94
CheshMITE	27.93	244	2.75E-09	102
FordPrefect	32.27	312	2.75E-09	118
FordPrefect_a	37.92	137	2.75E-09	138
Kanga1	31.04	350	2.75E-09	113
Kanga11a	33.32	303	2.75E-09	121
Kanga1a	31.12	248	2.75E-09	113
Kanga1b	30.75	169	2.75E-09	112
Kanga1c	30.91	368	2.75E-09	113
Kanga1d	29.00	252	2.75E-09	106
Kanga2_a	31.36	479	2.75E-09	114

Age of DNA transposons in cat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Looper	24.72	164	2.75E-09	90
MADE2	23.79	811	2.75E-09	87
MamRep1161	33.84	491	2.75E-09	123
MamRep137	36.04	658	2.75E-09	131
MamRep1879	32.92	521	2.75E-09	120
MamRep1894	34.64	168	2.75E-09	126
MamRep38	37.07	643	2.75E-09	135
MamRep4096	31.85	366	2.75E-09	116
MamRep434	33.51	642	2.75E-09	122
MARNA	36.65	1060	2.75E-09	133
MER102a	34.54	903	2.75E-09	126
MER102b	36.58	1190	2.75E-09	133
MER102c	36.17	1189	2.75E-09	132
MER103	30.10	2170	2.75E-09	110
MER104	29.60	702	2.75E-09	108
MER105	25.20	270	2.75E-09	92
MER106A	27.33	246	2.75E-09	100
MER106B	24.14	286	2.75E-09	88
MER112	29.23	1309	2.75E-09	106
MER113	30.76	939	2.75E-09	112
MER113A	30.26	524	2.75E-09	110
MER115	37.05	980	2.75E-09	135
MER117	35.58	1229	2.75E-09	130
MER119	28.19	484	2.75E-09	103
MER20	24.77	6001	2.75E-09	90
MER20B	33.33	1393	2.75E-09	121
MER3	25.05	4625	2.75E-09	91
MER33	23.36	3382	2.75E-09	85
MER44A	26.64	477	2.75E-09	97
MER44D	29.14	116	2.75E-09	106
MER45A	27.56	1160	2.75E-09	100
MER45B	26.49	554	2.75E-09	96
MER45C	33.40	503	2.75E-09	122
MER45R	27.31	386	2.75E-09	99
MER46C	31.24	1010	2.75E-09	114
MER53	20.86	2286	2.75E-09	76
MER58A	28.14	4352	2.75E-09	102
MER58B	26.58	2680	2.75E-09	97
MER58C	28.38	954	2.75E-09	103
MER58D	23.39	491	2.75E-09	85
MER5A	27.92	8860	2.75E-09	102
MER5A1	26.23	6226	2.75E-09	96
MER5B	30.39	6560	2.75E-09	111
MER5C	27.84	493	2.75E-09	101
MER5C1	30.40	226	2.75E-09	111
MER63A	31.84	1101	2.75E-09	116
MER63B	27.73	844	2.75E-09	101
MER63C	27.66	198	2.75E-09	101

Age of DNA transposons in cat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER63D	26.28	1054	2.75E-09	96
MER81	29.37	1229	2.75E-09	107
MER82	24.75	1328	2.75E-09	90
MER91A	37.55	638	2.75E-09	137
MER91B	31.58	390	2.75E-09	115
MER91C	30.01	297	2.75E-09	109
MER94	30.88	1461	2.75E-09	112
MER96	28.32	395	2.75E-09	103
MER96B	23.84	736	2.75E-09	87
MER97c	29.37	561	2.75E-09	107
MER99	27.34	122	2.75E-09	100
ORSL	23.71	396	2.75E-09	86
ORSL-2b	30.15	239	2.75E-09	110
Ricksha	36.28	209	2.75E-09	132
Ricksha_c	34.65	5752	2.75E-09	126
Tigger1	17.34	1946	2.75E-09	63
Tigger10	37.28	366	2.75E-09	136
Tigger11a	34.81	132	2.75E-09	127
Tigger12	37.00	156	2.75E-09	135
Tigger12A	35.61	189	2.75E-09	130
Tigger12c	37.26	424	2.75E-09	136
Tigger13a	33.56	1048	2.75E-09	122
Tigger14a	33.14	425	2.75E-09	121
Tigger15a	36.10	1469	2.75E-09	131
Tigger16a	33.48	270	2.75E-09	122
Tigger16b	35.07	329	2.75E-09	128
Tigger1a_Art	19.88	144	2.75E-09	72
Tigger1a_Car	17.80	1101	2.75E-09	65
Tigger1a_Mars	19.69	341	2.75E-09	72
Tigger2	20.25	240	2.75E-09	74
Tigger2a	21.05	209	2.75E-09	77
Tigger2a_Car	18.21	2876	2.75E-09	66
Tigger2b	20.93	179	2.75E-09	76
Tigger2b_Pri	19.39	399	2.75E-09	71
Tigger2f	18.12	830	2.75E-09	66
Tigger5	24.97	739	2.75E-09	91
Tigger5a	18.09	874	2.75E-09	66
Tigger6a	25.66	255	2.75E-09	93
Tigger6b	24.43	206	2.75E-09	89
Tigger7	17.76	1839	2.75E-09	65
Tigger8	36.92	250	2.75E-09	134
Tigger9a	31.77	355	2.75E-09	116
Tigger9b	27.17	378	2.75E-09	99
Zaphod	29.42	775	2.75E-09	107
Zaphod2	28.98	279	2.75E-09	106

Age of DNA transposons in cow

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	30.45	1689	2.80E-09	109
Arthur1A	30.94	607	2.80E-09	110
Arthur1B	32.11	1559	2.80E-09	115
Arthur1C	27.72	535	2.80E-09	99
BLACKJACK	30.83	1593	2.80E-09	110
Charlie1	25.43	3162	2.80E-09	91
Charlie10	30.16	913	2.80E-09	108
Charlie10a	32.07	200	2.80E-09	115
Charlie10b	31.04	193	2.80E-09	111
Charlie13a	34.56	365	2.80E-09	123
Charlie13b	34.78	208	2.80E-09	124
Charlie14a	31.71	336	2.80E-09	113
Charlie15a	35.76	1302	2.80E-09	128
Charlie16a	33.65	825	2.80E-09	120
Charlie17a	36.04	570	2.80E-09	129
Charlie18a	31.13	1275	2.80E-09	111
Charlie1a	27.28	2049	2.80E-09	97
Charlie1b	25.79	1472	2.80E-09	92
Charlie1b_Mars	24.50	1827	2.80E-09	88
Charlie2	31.63	914	2.80E-09	113
Charlie20a	34.07	467	2.80E-09	122
Charlie21a	31.79	414	2.80E-09	114
Charlie22a	33.76	515	2.80E-09	121
Charlie23a	33.92	675	2.80E-09	121
Charlie24	34.20	964	2.80E-09	122
Charlie25	36.22	699	2.80E-09	129
Charlie26a	31.06	291	2.80E-09	111
Charlie2a	30.14	788	2.80E-09	108
Charlie2b	29.96	2221	2.80E-09	107
Charlie3_Bt	16.68	4722	2.80E-09	60
Charlie4	29.22	166	2.80E-09	104
Charlie4a	27.53	2496	2.80E-09	98
Charlie4z	32.50	3042	2.80E-09	116
Charlie5	23.58	4379	2.80E-09	84
Charlie6	28.90	203	2.80E-09	103
Charlie7	32.45	2271	2.80E-09	116
Charlie7a	29.04	756	2.80E-09	104
Charlie8	34.69	2245	2.80E-09	124
Charlie9	29.41	997	2.80E-09	105
Cheshire	26.67	485	2.80E-09	95
Cheshire_Mars	26.26	1525	2.80E-09	94
CheshMITE	28.86	365	2.80E-09	103
FordPrefect	33.49	407	2.80E-09	120
FordPrefect_a	39.00	171	2.80E-09	139
HSMAR1_Bt	13.98	2648	2.80E-09	50
Kanga1	30.88	718	2.80E-09	110
Kanga11a	34.39	561	2.80E-09	123
Kanga1a	31.66	474	2.80E-09	113

Age of DNA transposons in cow

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1b	31.43	285	2.80E-09	112
Kanga1c	31.88	469	2.80E-09	114
Kanga1d	29.31	432	2.80E-09	105
Kanga2_a	31.52	999	2.80E-09	113
Looper	24.97	400	2.80E-09	89
MADE2	24.50	1360	2.80E-09	87
MamRep1161	33.69	975	2.80E-09	120
MamRep137	36.28	1128	2.80E-09	130
MamRep1879	33.74	646	2.80E-09	121
MamRep1894	36.88	233	2.80E-09	132
MamRep38	37.71	956	2.80E-09	135
MamRep4096	32.49	658	2.80E-09	116
MamRep434	33.69	1045	2.80E-09	120
MARNA	36.94	1726	2.80E-09	132
MER102a	34.77	1662	2.80E-09	124
MER102b	37.13	2210	2.80E-09	133
MER102c	36.58	2086	2.80E-09	131
MER103	30.42	3707	2.80E-09	109
MER104	28.66	1498	2.80E-09	102
MER105	26.02	495	2.80E-09	93
MER106A	28.31	475	2.80E-09	101
MER106B	25.73	608	2.80E-09	92
MER112	31.22	1967	2.80E-09	111
MER113	31.47	2168	2.80E-09	112
MER113A	30.71	981	2.80E-09	110
MER115	37.85	1520	2.80E-09	135
MER117	34.77	2288	2.80E-09	124
MER119	30.00	879	2.80E-09	107
MER1A_Bt_1	21.60	3986	2.80E-09	77
MER1A_Bt_2	18.94	2543	2.80E-09	68
MER2	23.36	1590	2.80E-09	83
MER20	26.71	12045	2.80E-09	95
MER20B	33.45	2507	2.80E-09	119
MER2B	27.04	4680	2.80E-09	97
MER3	26.74	8357	2.80E-09	96
MER33	24.56	4535	2.80E-09	88
MER44A	17.98	293	2.80E-09	64
MER44B	25.00	133	2.80E-09	89
MER44C	21.83	260	2.80E-09	78
MER45A	29.52	2207	2.80E-09	105
MER45B	27.27	971	2.80E-09	97
MER45C	34.53	759	2.80E-09	123
MER45R	27.27	656	2.80E-09	97
MER46C	31.70	1869	2.80E-09	113
MER47B	27.27	119	2.80E-09	97
MER53	21.82	4512	2.80E-09	78
MER58A	30.02	8158	2.80E-09	107
MER58B	28.09	5259	2.80E-09	100

Age of DNA transposons in cow

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER58C	29.59	1711	2.80E-09	106
MER58D	24.21	882	2.80E-09	86
MER5A	28.42	16236	2.80E-09	102
MER5A1	27.26	11126	2.80E-09	97
MER5B	29.97	9351	2.80E-09	107
MER5C	26.90	811	2.80E-09	96
MER5C1	31.15	479	2.80E-09	111
MER63A	32.83	1790	2.80E-09	117
MER63B	28.41	1588	2.80E-09	101
MER63C	28.08	452	2.80E-09	100
MER63D	26.90	2188	2.80E-09	96
MER8	34.46	515	2.80E-09	123
MER81	31.02	2093	2.80E-09	111
MER82	26.69	2247	2.80E-09	95
MER91A	37.92	950	2.80E-09	135
MER91B	31.30	647	2.80E-09	112
MER91C	29.77	476	2.80E-09	106
MER94	31.54	2310	2.80E-09	113
MER96	29.44	670	2.80E-09	105
MER96B	24.95	1419	2.80E-09	89
MER97a	29.01	105	2.80E-09	104
MER97b	30.35	128	2.80E-09	108
MER97c	30.48	1072	2.80E-09	109
MER99	27.56	237	2.80E-09	98
Oamar1	14.71	1153	2.80E-09	53
Oamar1_Na_1	17.58	8069	2.80E-09	63
ORSL	25.44	768	2.80E-09	91
ORSL-2a	31.41	173	2.80E-09	112
ORSL-2b	30.29	503	2.80E-09	108
Ricksha	37.58	529	2.80E-09	134
Ricksha_b	38.02	114	2.80E-09	136
Ricksha_c	35.19	8261	2.80E-09	126
Tigger1	18.06	3649	2.80E-09	64
Tigger10	37.19	548	2.80E-09	133
Tigger11a	34.78	221	2.80E-09	124
Tigger12	36.25	277	2.80E-09	129
Tigger12A	35.85	293	2.80E-09	128
Tigger12c	38.41	625	2.80E-09	137
Tigger13a	34.06	1863	2.80E-09	122
Tigger14a	32.96	643	2.80E-09	118
Tigger15a	36.27	2459	2.80E-09	130
Tigger16b	33.28	790	2.80E-09	119
Tigger1a_Art	20.06	2794	2.80E-09	72
Tigger1a_Car	21.04	1597	2.80E-09	75
Tigger1a_Mars	20.33	1387	2.80E-09	73
Tigger2	20.51	161	2.80E-09	73
Tigger2a_Art	29.28	4866	2.80E-09	105
Tigger2a_Car	21.39	729	2.80E-09	76

Age of DNA transposons in cow

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger2b	22.85	208	2.80E-09	82
Tigger2b_Pri	21.42	431	2.80E-09	76
Tigger2f	18.80	1152	2.80E-09	67
Tigger5	28.32	1202	2.80E-09	101
Tigger5a	19.28	1508	2.80E-09	69
Tigger6a	25.91	474	2.80E-09	93
Tigger6b	25.35	364	2.80E-09	91
Tigger7	21.90	1952	2.80E-09	78
Tigger8	36.57	460	2.80E-09	131
Tigger9a	32.95	633	2.80E-09	118
Tigger9b	28.26	621	2.80E-09	101
Zaphod	29.92	1514	2.80E-09	107
Zaphod2	29.91	439	2.80E-09	107

Age of DNA transposons in dog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	30.27	1948	2.74E-09	111
Arthur1A	30.60	824	2.74E-09	112
Arthur1b	30.99	2553	2.74E-09	113
Arthur1C	28.63	402	2.74E-09	105
BLACKJACK	30.68	2094	2.74E-09	112
Charlie1	25.85	1278	2.74E-09	94
Charlie10	29.43	1147	2.74E-09	108
Charlie10a	32.10	266	2.74E-09	117
Charlie10b	31.66	257	2.74E-09	116
Charlie11	34.40	113	2.74E-09	126
Charlie13a	34.91	486	2.74E-09	128
Charlie13b	34.98	263	2.74E-09	128
Charlie14a	31.02	470	2.74E-09	113
Charlie15a	36.19	1805	2.74E-09	132
Charlie16a	33.86	1137	2.74E-09	124
Charlie17a	36.17	806	2.74E-09	132
Charlie18a	31.24	1626	2.74E-09	114
Charlie19a	34.16	1001	2.74E-09	125
Charlie1a	25.39	4419	2.74E-09	93
Charlie1b	25.06	2383	2.74E-09	92
Charlie1b_Mars	23.61	2280	2.74E-09	86
Charlie2	30.53	1314	2.74E-09	112
Charlie20a	34.61	610	2.74E-09	126
Charlie21a	31.52	524	2.74E-09	115
Charlie22a	33.72	676	2.74E-09	123
Charlie23a	34.12	931	2.74E-09	125
Charlie24	33.99	1212	2.74E-09	124
Charlie25	36.06	923	2.74E-09	132
Charlie26a	31.01	378	2.74E-09	113
Charlie2a	30.69	1480	2.74E-09	112
Charlie2b	29.67	2328	2.74E-09	108
Charlie4	28.64	189	2.74E-09	105
Charlie4a	27.50	2861	2.74E-09	100
Charlie4z	32.36	3882	2.74E-09	118
Charlie5	23.76	2557	2.74E-09	87
Charlie6	28.47	217	2.74E-09	104
Charlie7	32.38	3142	2.74E-09	118
Charlie7a	30.00	989	2.74E-09	110
Charlie8	34.43	2798	2.74E-09	126
Charlie9	29.09	1095	2.74E-09	106
Cheshire	25.16	538	2.74E-09	92
Cheshire_Mars	25.89	1549	2.74E-09	95
CheshMITE	28.45	524	2.74E-09	104
FordPrefect	33.55	530	2.74E-09	123
FordPrefect_a	38.59	246	2.74E-09	141
Kanga1	30.92	838	2.74E-09	113
Kanga11a	34.16	683	2.74E-09	125
Kanga1a	31.39	599	2.74E-09	115

Age of DNA transposons in dog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1b	31.51	371	2.74E-09	115
Kanga1c	31.18	812	2.74E-09	114
Kanga1d	29.02	570	2.74E-09	106
Kanga2_a	31.79	1258	2.74E-09	116
Looper	24.38	511	2.74E-09	89
MADE2	23.86	1809	2.74E-09	87
MamRep1161	34.13	1050	2.74E-09	125
MamRep137	36.57	1517	2.74E-09	134
MamRep1879	33.66	1100	2.74E-09	123
MamRep1894	36.75	345	2.74E-09	134
MamRep38	37.59	1288	2.74E-09	137
MamRep4096	32.45	803	2.74E-09	119
MamRep434	33.68	1390	2.74E-09	123
MARNA	36.77	2284	2.74E-09	134
MER102a	34.94	2234	2.74E-09	128
MER102b	37.15	2848	2.74E-09	136
MER102c	36.44	2653	2.74E-09	133
MER103	30.34	4826	2.74E-09	111
MER104	29.99	1646	2.74E-09	110
MER105	25.48	641	2.74E-09	93
MER106A	27.95	659	2.74E-09	102
MER106B	25.64	742	2.74E-09	94
MER112	29.73	2794	2.74E-09	109
MER113	31.51	2553	2.74E-09	115
MER113A	31.00	1237	2.74E-09	113
MER115	38.25	2011	2.74E-09	140
MER117	35.86	2811	2.74E-09	131
MER119	29.52	1081	2.74E-09	108
MER20	25.48	14431	2.74E-09	93
MER20B	33.60	3302	2.74E-09	123
MER3	25.85	10793	2.74E-09	94
MER33	23.53	8125	2.74E-09	86
MER44A	28.28	1045	2.74E-09	103
MER44D	27.01	272	2.74E-09	99
MER45A	28.81	2727	2.74E-09	105
MER45B	27.33	1242	2.74E-09	100
MER45C	34.78	1070	2.74E-09	127
MER45R	26.95	867	2.74E-09	98
MER46C	31.40	2296	2.74E-09	115
MER47B	25.50	167	2.74E-09	93
MER53	21.25	5430	2.74E-09	78
MER58A	29.09	10578	2.74E-09	106
MER58B	27.27	6020	2.74E-09	100
MER58C	29.09	2250	2.74E-09	106
MER58D	24.04	1050	2.74E-09	88
MER5A	28.67	19270	2.74E-09	105
MER5A1	26.87	13609	2.74E-09	98
MER5B	31.06	13820	2.74E-09	113

Age of DNA transposons in dog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER5C	27.91	1079	2.74E-09	102
MER5C1	31.59	486	2.74E-09	115
MER63A	32.25	2620	2.74E-09	118
MER63B	27.81	1903	2.74E-09	102
MER63C	27.47	540	2.74E-09	100
MER63D	26.44	2309	2.74E-09	97
MER81	30.54	2551	2.74E-09	112
MER82	25.67	2922	2.74E-09	94
MER91A	38.38	1276	2.74E-09	140
MER91B	31.61	877	2.74E-09	115
MER91C	30.28	663	2.74E-09	111
MER94	31.57	3131	2.74E-09	115
MER96	29.56	856	2.74E-09	108
MER96B	24.47	1800	2.74E-09	89
MER97a	28.91	137	2.74E-09	106
MER97b	30.11	144	2.74E-09	110
MER97c	30.16	1311	2.74E-09	110
MER99	27.33	259	2.74E-09	100
ORSL	24.83	947	2.74E-09	91
ORSL-2a	31.00	246	2.74E-09	113
ORSL-2b	30.19	545	2.74E-09	110
Ricksha	36.88	831	2.74E-09	135
Ricksha_b	37.42	168	2.74E-09	137
Ricksha_c	35.29	12455	2.74E-09	129
Tigger1	17.04	3852	2.74E-09	62
Tigger10	37.37	768	2.74E-09	137
Tigger11a	35.48	309	2.74E-09	130
Tigger12	35.94	357	2.74E-09	131
Tigger12A	35.71	377	2.74E-09	130
Tigger12c	37.79	890	2.74E-09	138
Tigger13a	34.10	2345	2.74E-09	125
Tigger14a	33.34	870	2.74E-09	122
Tigger15a	36.22	3149	2.74E-09	132
Tigger16a	34.24	563	2.74E-09	125
Tigger16b	35.64	655	2.74E-09	130
Tigger1a_Art	20.70	301	2.74E-09	76
Tigger1a_Car	18.44	2327	2.74E-09	67
Tigger1a_Mars	21.21	627	2.74E-09	77
Tigger2	18.89	462	2.74E-09	69
Tigger2a	21.11	475	2.74E-09	77
Tigger2a_Art	23.29	174	2.74E-09	85
Tigger2a_Car	19.28	7531	2.74E-09	70
Tigger2b	22.39	322	2.74E-09	82
Tigger2b_Pri	20.03	818	2.74E-09	73
Tigger2f	17.88	1713	2.74E-09	65
Tigger5	25.64	1719	2.74E-09	94
Tigger5a	18.69	1952	2.74E-09	68
Tigger6a	25.43	541	2.74E-09	93

Age of DNA transposons in dog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger6b	24.28	442	2.74E-09	89
Tigger7	18.55	4021	2.74E-09	68
Tigger8	36.36	592	2.74E-09	133
Tigger9a	32.62	789	2.74E-09	119
Tigger9b	27.99	763	2.74E-09	102
Zaphod	30.09	1884	2.74E-09	110
Zaphod2	29.11	634	2.74E-09	106

Age of DNA transposons in elephant

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	27.24	1298	2.57E-09	106
arthur1a	27.06	175	2.57E-09	106
arthur1b	29.20	603	2.57E-09	114
Arthur1C	24.92	358	2.57E-09	97
BLACKJACK	28.19	1793	2.57E-09	110
Charlie1	22.67	953	2.57E-09	88
Charlie10	27.14	918	2.57E-09	106
Charlie10a	29.34	231	2.57E-09	114
Charlie10b	28.11	218	2.57E-09	110
Charlie13a	32.89	408	2.57E-09	128
Charlie13b	34.15	252	2.57E-09	133
Charlie14a	29.62	535	2.57E-09	115
Charlie15a	35.31	2285	2.57E-09	138
Charlie16a	33.36	1227	2.57E-09	130
Charlie17a	33.68	917	2.57E-09	131
Charlie18a	29.57	1539	2.57E-09	115
Charlie19a	33.38	1060	2.57E-09	130
Charlie1a	22.69	3118	2.57E-09	88
Charlie1b	21.91	1718	2.57E-09	85
Charlie1b_Mars	21.01	1433	2.57E-09	82
Charlie2	28.86	932	2.57E-09	113
Charlie20a	32.37	566	2.57E-09	126
Charlie21a	30.11	488	2.57E-09	117
Charlie22a	32.44	687	2.57E-09	126
Charlie23a	32.98	919	2.57E-09	129
Charlie24	32.17	1078	2.57E-09	125
Charlie25	34.28	727	2.57E-09	134
Charlie26a	29.52	335	2.57E-09	115
Charlie2a	28.92	1174	2.57E-09	113
Charlie2b	27.67	1894	2.57E-09	108
Charlie4	27.63	117	2.57E-09	108
Charlie4a	24.50	2396	2.57E-09	95
Charlie4z	31.00	4114	2.57E-09	121
Charlie5	20.66	1867	2.57E-09	81
Charlie6	25.74	175	2.57E-09	100
Charlie7	30.55	2687	2.57E-09	119
Charlie7a	27.91	1042	2.57E-09	109
Charlie8	32.94	2174	2.57E-09	128
Charlie9	28.17	1044	2.57E-09	110
Cheshire	22.88	317	2.57E-09	89
Cheshire_Mars	22.42	1005	2.57E-09	87
CheshMITE	26.06	351	2.57E-09	102
Dumbo	15.96	2919	2.57E-09	62
Dumbo_NA_1	21.37	4149	2.57E-09	83
Dumbo_NA_2	15.66	2762	2.57E-09	61
FordPrefect	31.02	433	2.57E-09	121
FordPrefect_a	35.42	270	2.57E-09	138
Kanga1	28.27	588	2.57E-09	110

Age of DNA transposons in elephant

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga11a	32.64	659	2.57E-09	127
Kanga1a	29.62	499	2.57E-09	115
Kanga1b	28.77	328	2.57E-09	112
Kanga1c	28.70	689	2.57E-09	112
Kanga1d	26.60	431	2.57E-09	104
Kanga2_a	29.66	1016	2.57E-09	116
LAMAR2_La	11.57	638	2.57E-09	45
Looper	21.48	343	2.57E-09	84
MADE2	22.62	1852	2.57E-09	88
MamRep1161	32.90	1215	2.57E-09	128
MamRep137	34.95	1504	2.57E-09	136
MamRep1879	31.61	751	2.57E-09	123
MamRep1894	35.18	423	2.57E-09	137
MamRep38	34.89	1324	2.57E-09	136
MamRep4096	30.03	802	2.57E-09	117
MamRep434	33.41	1524	2.57E-09	130
MARNA	37.27	2773	2.57E-09	145
MER102a	33.12	1917	2.57E-09	129
MER102b	35.43	2787	2.57E-09	138
MER102c	34.22	2805	2.57E-09	133
MER103	29.06	4731	2.57E-09	113
MER104	27.89	1390	2.57E-09	109
MER105	23.15	562	2.57E-09	90
MER106A	24.60	532	2.57E-09	96
MER106B	21.53	594	2.57E-09	84
MER112	27.80	2671	2.57E-09	108
MER113	29.01	2471	2.57E-09	113
MER113A	28.97	1283	2.57E-09	113
MER115	35.39	1650	2.57E-09	138
MER117	34.46	2963	2.57E-09	134
MER119	25.71	911	2.57E-09	100
MER2	15.82	1568	2.57E-09	62
MER20	21.11	11689	2.57E-09	82
MER20B	31.02	2926	2.57E-09	121
MER2B	22.50	1357	2.57E-09	88
MER3	22.34	8975	2.57E-09	87
MER33	20.09	6455	2.57E-09	78
MER44A	19.21	422	2.57E-09	75
MER45A	24.50	2108	2.57E-09	96
MER45B	23.00	962	2.57E-09	90
MER45C	31.48	761	2.57E-09	123
MER45R	24.40	631	2.57E-09	95
MER46C	29.19	1890	2.57E-09	114
MER53	17.99	4423	2.57E-09	70
MER58A	25.21	7958	2.57E-09	98
MER58B	23.23	4967	2.57E-09	91
MER58C	26.29	1710	2.57E-09	102
MER58D	20.64	1052	2.57E-09	80

Age of DNA transposons in elephant

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER5A	25.54	21306	2.57E-09	100
MER5A1	23.00	17050	2.57E-09	90
MER5B	28.89	12570	2.57E-09	113
MER5C	25.40	924	2.57E-09	99
MER5C1	28.73	452	2.57E-09	112
MER63A	29.28	2125	2.57E-09	114
MER63B	23.97	1479	2.57E-09	93
MER63C	24.47	396	2.57E-09	95
MER63D	23.10	2019	2.57E-09	90
MER81	26.68	2247	2.57E-09	104
MER82	21.08	2193	2.57E-09	82
MER91A	36.59	1328	2.57E-09	143
MER91B	30.02	795	2.57E-09	117
MER91C	28.88	647	2.57E-09	113
MER94	29.16	3008	2.57E-09	114
MER96	24.28	932	2.57E-09	95
MER96B	20.68	1375	2.57E-09	81
MER97a	25.97	101	2.57E-09	101
MER97c	26.65	1074	2.57E-09	104
MER99	23.59	193	2.57E-09	92
ORSL	21.02	748	2.57E-09	82
ORSL-2a	28.90	204	2.57E-09	113
ORSL-2b	28.14	500	2.57E-09	110
Ricksha	32.91	1133	2.57E-09	128
Ricksha_c	30.04	15866	2.57E-09	117
Tigger1	13.23	3359	2.57E-09	52
Tigger10	36.45	689	2.57E-09	142
Tigger11a	34.91	311	2.57E-09	136
Tigger12	35.05	348	2.57E-09	137
Tigger12A	34.98	409	2.57E-09	136
Tigger12c	37.11	923	2.57E-09	145
Tigger13a	32.04	2106	2.57E-09	125
Tigger14a	32.51	868	2.57E-09	127
Tigger15a	35.49	3318	2.57E-09	138
Tigger16a	31.99	1222	2.57E-09	125
Tigger16b	33.63	1204	2.57E-09	131
Tigger1a_Art	14.10	181	2.57E-09	55
Tigger1a_Car	14.95	118	2.57E-09	58
Tigger1a_Mars	13.23	269	2.57E-09	52
Tigger3(Golem)	35.63	3101	2.57E-09	139
Tigger3b	36.41	132	2.57E-09	142
Tigger5	43.50	191	2.57E-09	170
Tigger6a	21.49	473	2.57E-09	84
Tigger6b	21.27	374	2.57E-09	83
Tigger7	19.50	730	2.57E-09	76
Tigger8	35.18	571	2.57E-09	137
Tigger9a	29.94	673	2.57E-09	117
Tigger9b	23.67	557	2.57E-09	92

Age of DNA transposons in elephant

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Zaphod	27.73	1383	2.57E-09	108
Zaphod2	27.73	539	2.57E-09	108

Age of DNA transposons in guinea pig

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	34.02	822	3.2796E-09	104
Arthur1A	33.94	217	3.2796E-09	103
Arthur1B	34.88	981	3.2796E-09	106
Arthur1C	32.86	172	3.2796E-09	100
BLACKJACK	34.36	767	3.2796E-09	105
Charlie1	32.01	566	3.2796E-09	98
Charlie10	33.53	485	3.2796E-09	102
Charlie12	31.35	299	3.2796E-09	96
Charlie13a	36.04	213	3.2796E-09	110
Charlie14a	30.98	102	3.2796E-09	94
Charlie15a	35.48	285	3.2796E-09	108
Charlie16a	33.57	220	3.2796E-09	102
Charlie17a	37.44	103	3.2796E-09	114
Charlie18a	32.21	366	3.2796E-09	98
Charlie19a	34.34	260	3.2796E-09	105
Charlie1a	31.57	2067	3.2796E-09	96
Charlie1b	30.20	975	3.2796E-09	92
Charlie1b_Mars	29.82	1616	3.2796E-09	91
Charlie2	32.45	479	3.2796E-09	99
Charlie20a	35.28	191	3.2796E-09	108
Charlie21a	33.10	179	3.2796E-09	101
Charlie22a	33.75	193	3.2796E-09	103
Charlie23a	33.28	175	3.2796E-09	101
Charlie24	36.26	370	3.2796E-09	111
Charlie25	38.82	343	3.2796E-09	118
Charlie26a	32.57	104	3.2796E-09	99
Charlie2a	33.37	497	3.2796E-09	102
Charlie2b	32.39	791	3.2796E-09	99
Charlie3	25.41	1808	3.2796E-09	77
Charlie4	30.92	118	3.2796E-09	94
Charlie4a	31.25	1323	3.2796E-09	95
Charlie4z	33.38	712	3.2796E-09	102
Charlie5	28.99	1404	3.2796E-09	88
Charlie6	34.16	126	3.2796E-09	104
Charlie7	34.29	882	3.2796E-09	105
Charlie7a	29.96	162	3.2796E-09	91
Charlie8	34.79	791	3.2796E-09	106
Charlie9	31.39	300	3.2796E-09	96
Cheshire	31.28	363	3.2796E-09	95
Cheshire_Mars	31.17	1319	3.2796E-09	95
CheshMITE	31.01	153	3.2796E-09	95
FordPrefect	38.31	273	3.2796E-09	117
Kanga1	34.32	320	3.2796E-09	105
Kanga11a	36.19	214	3.2796E-09	110
Kanga1a	35.44	217	3.2796E-09	108
Kanga1b	35.60	124	3.2796E-09	109
Kanga1c	35.13	293	3.2796E-09	107
Kanga1d	30.84	208	3.2796E-09	94

Age of DNA transposons in guinea pig

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga2_a	34.28	398	3.2796E-09	105
Looper	29.97	219	3.2796E-09	91
MADE2	25.12	381	3.2796E-09	77
MamRep1161	34.26	316	3.2796E-09	104
MamRep137	38.03	365	3.2796E-09	116
MamRep1879	34.12	174	3.2796E-09	104
MamRep38	40.25	230	3.2796E-09	123
MamRep4096	33.89	227	3.2796E-09	103
MamRep434	33.36	449	3.2796E-09	102
MARNA	35.89	1982	3.2796E-09	109
MER102a	35.70	596	3.2796E-09	109
MER102b	37.99	624	3.2796E-09	116
MER102c	38.05	477	3.2796E-09	116
MER103	31.58	1069	3.2796E-09	96
MER104	32.43	436	3.2796E-09	99
MER105	30.72	231	3.2796E-09	94
MER106A	33.12	158	3.2796E-09	101
MER106B	31.10	236	3.2796E-09	95
MER112	31.22	599	3.2796E-09	95
MER113	32.74	788	3.2796E-09	100
MER113A	31.91	257	3.2796E-09	97
MER115	40.05	510	3.2796E-09	122
MER117	35.82	728	3.2796E-09	109
MER119	34.74	460	3.2796E-09	106
MER1A	30.37	1174	3.2796E-09	93
MER1B	30.16	550	3.2796E-09	92
MER2	30.67	8364	3.2796E-09	94
MER20	32.84	6500	3.2796E-09	100
MER20B	36.31	1033	3.2796E-09	111
MER2B	33.74	1338	3.2796E-09	103
MER3	31.28	3709	3.2796E-09	95
MER30	0.00	0	3.2796E-09	0
MER30B	0.00	0	3.2796E-09	0
MER33	28.29	3238	3.2796E-09	86
MER44A	29.76	859	3.2796E-09	91
MER44B	33.24	988	3.2796E-09	101
MER44C	32.90	136	3.2796E-09	100
MER44D	33.76	730	3.2796E-09	103
MER45A	35.16	970	3.2796E-09	107
MER45B	34.01	685	3.2796E-09	104
MER45C	38.08	366	3.2796E-09	116
MER45R	32.32	401	3.2796E-09	99
MER46C	34.46	804	3.2796E-09	105
MER47B	36.17	409	3.2796E-09	110
MER53	26.39	2345	3.2796E-09	80
MER58A	35.49	4107	3.2796E-09	108
MER58B	33.06	2919	3.2796E-09	101
MER58C	31.97	646	3.2796E-09	97

Age of DNA transposons in guinea pig

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER58D	28.03	402	3.2796E-09	85
MER5A	33.57	6416	3.2796E-09	102
MER5A1	31.71	4836	3.2796E-09	97
MER5B	34.16	3718	3.2796E-09	104
MER5C	30.86	374	3.2796E-09	94
MER5C1	33.57	142	3.2796E-09	102
MER63A	37.56	676	3.2796E-09	115
MER63B	33.53	787	3.2796E-09	102
MER63C	33.79	287	3.2796E-09	103
MER63D	31.43	1242	3.2796E-09	96
MER81	35.06	681	3.2796E-09	107
MER82	33.66	1535	3.2796E-09	103
MER91A	40.07	212	3.2796E-09	122
MER91B	33.58	234	3.2796E-09	102
MER94	33.51	594	3.2796E-09	102
MER96	33.42	150	3.2796E-09	102
MER96B	29.38	652	3.2796E-09	90
MER97c	34.50	478	3.2796E-09	105
MER99	32.29	129	3.2796E-09	98
ORSL	28.08	296	3.2796E-09	86
ORSL-2b	31.91	245	3.2796E-09	97
Ricksha	33.23	608	3.2796E-09	101
Ricksha_b	35.94	153	3.2796E-09	110
Ricksha_c	35.85	8375	3.2796E-09	109
RMER30	30.84	483	3.2796E-09	94
Tigger1	30.16	830	3.2796E-09	92
Tigger10	37.72	192	3.2796E-09	115
Tigger12	39.94	156	3.2796E-09	122
Tigger12c	38.11	185	3.2796E-09	116
Tigger13a	35.62	759	3.2796E-09	109
Tigger14a	33.74	212	3.2796E-09	103
Tigger15a	36.42	693	3.2796E-09	111
Tigger16a	33.05	193	3.2796E-09	101
Tigger16b	35.21	228	3.2796E-09	107
Tigger1a_Mars	29.22	125	3.2796E-09	89
Tigger2b	30.02	169	3.2796E-09	92
Tigger3(Golem)	31.51	4630	3.2796E-09	96
Tigger3b	31.65	695	3.2796E-09	96
Tigger4(Zombi)	25.87	8435	3.2796E-09	79
Tigger5	35.43	2551	3.2796E-09	108
Tigger5a	29.68	2511	3.2796E-09	91
Tigger6a	33.26	314	3.2796E-09	101
Tigger6b	31.12	286	3.2796E-09	95
Tigger7	31.07	6097	3.2796E-09	95
Tigger8	36.66	163	3.2796E-09	112
Tigger9a	36.48	254	3.2796E-09	111
Tigger9b	33.27	329	3.2796E-09	101
Zaphod	33.59	724	3.2796E-09	102

Age of DNA transposons in guinea pig

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Zaphod2	31.83	166	3.2796E-09	97

Age of DNA transposons in hedgehog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	35.22	401	3.4996E-09	101
Arthur1A	33.21	106	3.4996E-09	95
Arthur1B	35.50	521	3.4996E-09	101
Arthur1C	33.77	104	3.4996E-09	97
BLACKJACK	33.72	375	3.4996E-09	96
Charlie1	32.74	286	3.4996E-09	94
Charlie10	33.56	229	3.4996E-09	96
Charlie13a	36.03	132	3.4996E-09	103
Charlie15a	33.75	133	3.4996E-09	96
Charlie16a	32.95	117	3.4996E-09	94
Charlie18a	32.19	157	3.4996E-09	92
Charlie19a	33.56	147	3.4996E-09	96
Charlie1a	31.59	1211	3.4996E-09	90
Charlie1b	30.86	556	3.4996E-09	88
Charlie1b_Mars	30.22	1171	3.4996E-09	86
Charlie2	32.66	266	3.4996E-09	93
Charlie20a	35.07	107	3.4996E-09	100
Charlie22a	32.78	113	3.4996E-09	94
Charlie24	36.18	241	3.4996E-09	103
Charlie25	38.01	217	3.4996E-09	109
Charlie2a	33.72	271	3.4996E-09	96
Charlie2b	31.96	432	3.4996E-09	91
Charlie4a	31.89	753	3.4996E-09	91
Charlie4z	33.38	346	3.4996E-09	95
Charlie5	29.53	753	3.4996E-09	84
Charlie7	34.30	474	3.4996E-09	98
Charlie8	34.88	436	3.4996E-09	100
Charlie9	31.27	148	3.4996E-09	89
Cheshire	32.69	222	3.4996E-09	93
Cheshire_Mars	31.36	930	3.4996E-09	90
FordPrefect	36.74	174	3.4996E-09	105
Kanga1	35.39	172	3.4996E-09	101
Kanga11a	35.20	118	3.4996E-09	101
Kanga1a	35.47	123	3.4996E-09	101
Kanga1c	34.74	161	3.4996E-09	99
Kanga2_a	33.15	225	3.4996E-09	95
Looper	29.79	163	3.4996E-09	85
MADE2	25.55	191	3.4996E-09	73
MamRep1161	33.88	203	3.4996E-09	97
MamRep137	37.66	226	3.4996E-09	108
MamRep1879	35.03	113	3.4996E-09	100
MamRep38	41.65	103	3.4996E-09	119
MamRep4096	33.95	123	3.4996E-09	97
MamRep434	33.10	357	3.4996E-09	95
MARNA	36.67	384	3.4996E-09	105
MER102a	35.47	269	3.4996E-09	101
MER102b	36.76	353	3.4996E-09	105
MER102c	37.51	326	3.4996E-09	107

Age of DNA transposons in hedgehog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER103	31.08	597	3.4996E-09	89
MER104	32.02	261	3.4996E-09	92
MER105	30.06	167	3.4996E-09	86
MER106B	30.91	144	3.4996E-09	88
MER112	29.75	276	3.4996E-09	85
MER113	32.79	415	3.4996E-09	94
MER113A	31.78	120	3.4996E-09	91
MER115	39.13	247	3.4996E-09	112
MER117	34.54	480	3.4996E-09	99
MER119	34.54	270	3.4996E-09	99
MER2	30.99	356	3.4996E-09	89
MER20	33.05	4082	3.4996E-09	94
MER20B	36.07	500	3.4996E-09	103
MER3	32.23	2046	3.4996E-09	92
MER33	29.54	1820	3.4996E-09	84
MER44A	29.55	532	3.4996E-09	84
MER45A	35.20	542	3.4996E-09	101
MER45B	33.98	406	3.4996E-09	97
MER45C	38.98	222	3.4996E-09	111
MER45R	33.12	272	3.4996E-09	95
MER46C	34.03	434	3.4996E-09	97
MER53	27.27	1350	3.4996E-09	78
MER58A	36.01	2189	3.4996E-09	103
MER58B	33.75	1877	3.4996E-09	96
MER58C	32.81	294	3.4996E-09	94
MER58D	27.94	221	3.4996E-09	80
MER5A	33.03	4389	3.4996E-09	94
MER5A1	31.48	3297	3.4996E-09	90
MER5B	33.42	2171	3.4996E-09	95
MER5C	31.16	191	3.4996E-09	89
MER6	34.58	566	3.4996E-09	99
MER6_Ee_1	31.97	1003	3.4996E-09	91
MER63A	38.07	351	3.4996E-09	109
MER63B	34.36	474	3.4996E-09	98
MER63C	33.63	150	3.4996E-09	96
MER63D	32.47	776	3.4996E-09	93
MER6A	30.83	1021	3.4996E-09	88
MER81	35.11	384	3.4996E-09	100
MER82	33.61	983	3.4996E-09	96
MER91A	39.46	119	3.4996E-09	113
MER91B	32.35	118	3.4996E-09	92
MER94	33.68	372	3.4996E-09	96
MER96B	29.13	284	3.4996E-09	83
MER97c	34.68	256	3.4996E-09	99
ORSL	29.94	151	3.4996E-09	86
Ricksha	34.33	107	3.4996E-09	98
Ricksha_c	37.35	1189	3.4996E-09	107
Tigger1	28.35	1400	3.4996E-09	81

Age of DNA transposons in hedgehog

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger10	36.55	136	3.4996E-09	104
Tigger12c	36.97	145	3.4996E-09	106
Tigger13a	35.15	423	3.4996E-09	100
Tigger14a	32.35	149	3.4996E-09	92
Tigger15a	35.75	441	3.4996E-09	102
Tigger16b	34.86	114	3.4996E-09	100
Tigger1a_Art	28.36	297	3.4996E-09	81
Tigger1a_Car	28.64	220	3.4996E-09	82
Tigger1a_Mars	28.46	588	3.4996E-09	81
Tigger2a	29.32	188	3.4996E-09	84
Tigger2a_Art	32.66	167	3.4996E-09	93
Tigger2a_Car	27.56	348	3.4996E-09	79
Tigger2b	28.88	184	3.4996E-09	83
Tigger2b_Pri	28.67	168	3.4996E-09	82
Tigger2f	28.09	633	3.4996E-09	80
Tigger3(Golem)	31.89	1721	3.4996E-09	91
Tigger5	32.21	444	3.4996E-09	92
Tigger5a	29.69	545	3.4996E-09	85
Tigger6a	32.43	193	3.4996E-09	93
Tigger6b	31.85	218	3.4996E-09	91
Tigger7	29.06	732	3.4996E-09	83
Tigger8	36.26	105	3.4996E-09	104
Tigger9a	34.99	140	3.4996E-09	100
Tigger9b	35.40	218	3.4996E-09	101
Zaphod	34.41	424	3.4996E-09	98
Zaphod2	31.73	124	3.4996E-09	91

Age of DNA transposons in horse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	27.06	1790	2.4804E-09	109
Arthur1a	27.60	1061	2.4804E-09	111
Arthur1b	28.42	2531	2.4804E-09	115
Arthur1C	25.57	431	2.4804E-09	103
BLACKJACK	28.40	2039	2.4804E-09	114
Charlie1	22.27	1253	2.4804E-09	90
Charlie10	27.07	1137	2.4804E-09	109
Charlie10a	29.94	259	2.4804E-09	121
Charlie10b	28.69	294	2.4804E-09	116
Charlie11	33.06	109	2.4804E-09	133
Charlie13a	33.75	499	2.4804E-09	136
Charlie13b	35.06	343	2.4804E-09	141
Charlie14a	30.80	695	2.4804E-09	124
Charlie15a	35.72	3159	2.4804E-09	144
Charlie16a	33.73	1714	2.4804E-09	136
Charlie17a	34.23	1254	2.4804E-09	138
Charlie18a	30.13	2199	2.4804E-09	121
Charlie19a	33.61	1421	2.4804E-09	136
Charlie1a	22.02	3787	2.4804E-09	89
Charlie1b	21.50	2245	2.4804E-09	87
Charlie1b_Mars	21.25	1295	2.4804E-09	86
Charlie2	29.14	1188	2.4804E-09	117
Charlie20a	33.35	777	2.4804E-09	134
Charlie21a	30.50	597	2.4804E-09	123
Charlie22a	33.00	964	2.4804E-09	133
Charlie23a	33.56	1280	2.4804E-09	135
Charlie24	32.63	1378	2.4804E-09	132
Charlie25	34.64	978	2.4804E-09	140
Charlie26a	30.39	444	2.4804E-09	123
Charlie2a	28.95	1601	2.4804E-09	117
Charlie2b	27.29	2400	2.4804E-09	110
Charlie4	27.69	137	2.4804E-09	112
Charlie4a	24.64	2671	2.4804E-09	99
Charlie4z	31.29	5726	2.4804E-09	126
Charlie5	20.20	1996	2.4804E-09	81
Charlie6	27.34	207	2.4804E-09	110
Charlie7	30.74	3392	2.4804E-09	124
Charlie7a	28.41	1544	2.4804E-09	115
Charlie8	33.46	2942	2.4804E-09	135
Charlie9	28.35	1239	2.4804E-09	114
Cheshire	21.55	390	2.4804E-09	87
Cheshire_Mars	21.82	882	2.4804E-09	88
CheshMITE	26.32	363	2.4804E-09	106
FordPrefect	29.20	439	2.4804E-09	118
FordPrefect_a	34.31	266	2.4804E-09	138
Kanga1	28.83	845	2.4804E-09	116
Kanga11a	32.99	855	2.4804E-09	133
Kanga1a	28.47	657	2.4804E-09	115

Age of DNA transposons in horse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1b	29.02	408	2.4804E-09	117
Kanga1c	29.01	863	2.4804E-09	117
Kanga1d	26.99	550	2.4804E-09	109
Kanga2_a	29.79	1370	2.4804E-09	120
Looper	20.70	443	2.4804E-09	83
MADE2	22.20	2570	2.4804E-09	90
MamRep1161	33.86	1418	2.4804E-09	137
MamRep137	35.23	2026	2.4804E-09	142
MamRep1879	31.95	1631	2.4804E-09	129
MamRep1894	34.87	652	2.4804E-09	141
MamRep38	35.21	1772	2.4804E-09	142
MamRep4096	30.91	1065	2.4804E-09	125
MamRep434	33.91	1879	2.4804E-09	137
MARNA	36.69	3012	2.4804E-09	148
MER102a	33.52	2537	2.4804E-09	135
MER102b	35.59	4118	2.4804E-09	143
MER102c	34.68	3870	2.4804E-09	140
MER103	29.51	6474	2.4804E-09	119
MER104	28.05	2149	2.4804E-09	113
MER105	23.03	648	2.4804E-09	93
MER106A	24.68	795	2.4804E-09	100
MER106B	21.60	707	2.4804E-09	87
MER112	28.29	3927	2.4804E-09	114
MER113	30.07	3164	2.4804E-09	121
MER113A	29.98	1799	2.4804E-09	121
MER115	35.19	2283	2.4804E-09	142
MER117	34.79	3953	2.4804E-09	140
MER119	25.50	1006	2.4804E-09	103
MER2	18.16	2079	2.4804E-09	73
MER20	21.09	14162	2.4804E-09	85
MER20B	31.61	3546	2.4804E-09	127
MER2B	20.80	694	2.4804E-09	84
MER3	22.17	11276	2.4804E-09	89
MER33	20.05	7503	2.4804E-09	81
MER44A	19.19	314	2.4804E-09	77
MER44B	18.60	441	2.4804E-09	75
MER44D	25.05	1231	2.4804E-09	101
MER45A	24.22	2903	2.4804E-09	98
MER45B	22.78	1030	2.4804E-09	92
MER45C	30.96	1075	2.4804E-09	125
MER45R	23.71	664	2.4804E-09	96
MER46C	29.37	2415	2.4804E-09	118
MER47B	18.97	123	2.4804E-09	76
MER53	18.07	5394	2.4804E-09	73
MER58A	25.29	10948	2.4804E-09	102
MER58B	23.61	5472	2.4804E-09	95
MER58C	26.93	2406	2.4804E-09	109
MER58D	20.41	1027	2.4804E-09	82

Age of DNA transposons in horse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER5A	26.66	22158	2.4804E-09	107
MER5A1	24.24	14520	2.4804E-09	98
MER5B	29.53	16926	2.4804E-09	119
MER5C	25.59	1154	2.4804E-09	103
MER5C1	29.29	608	2.4804E-09	118
MER63A	28.92	3184	2.4804E-09	117
MER63B	23.84	1795	2.4804E-09	96
MER63C	23.62	552	2.4804E-09	95
MER63D	23.09	1885	2.4804E-09	93
MER81	26.51	3208	2.4804E-09	107
MER82	21.20	2425	2.4804E-09	85
MER91A	35.97	1922	2.4804E-09	145
MER91B	30.23	1155	2.4804E-09	122
MER91C	29.15	961	2.4804E-09	118
MER94	29.60	4179	2.4804E-09	119
MER96	24.92	1165	2.4804E-09	100
MER96B	20.90	1718	2.4804E-09	84
MER97a	26.07	167	2.4804E-09	105
MER97b	26.34	135	2.4804E-09	106
MER97c	27.56	1319	2.4804E-09	111
MER99	23.10	249	2.4804E-09	93
ORSL	21.74	978	2.4804E-09	88
ORSL-2a	29.84	305	2.4804E-09	120
ORSL-2b	28.78	589	2.4804E-09	116
Ricksha	33.99	904	2.4804E-09	137
Ricksha_b	36.88	208	2.4804E-09	149
Ricksha_c	32.74	16090	2.4804E-09	132
Tigger1	11.61	2491	2.4804E-09	47
Tigger10	37.28	1001	2.4804E-09	150
Tigger11a	34.71	407	2.4804E-09	140
Tigger12	35.31	488	2.4804E-09	142
Tigger12A	35.91	529	2.4804E-09	145
Tigger12c	37.32	1169	2.4804E-09	150
Tigger13a	32.39	2657	2.4804E-09	131
Tigger14a	33.28	1243	2.4804E-09	134
Tigger15a	36.11	4637	2.4804E-09	146
Tigger16a	33.96	815	2.4804E-09	137
Tigger16b	35.56	951	2.4804E-09	143
Tigger1a_Art	15.08	263	2.4804E-09	61
Tigger1a_Car	16.15	573	2.4804E-09	65
Tigger1a_Mars	12.73	924	2.4804E-09	51
Tigger2a_Art	23.79	165	2.4804E-09	96
Tigger3(Golem)	15.42	23698	2.4804E-09	62
Tigger3b	14.66	452	2.4804E-09	59
Tigger5	19.51	1154	2.4804E-09	79
Tigger5a	13.66	1247	2.4804E-09	55
Tigger6a	21.10	493	2.4804E-09	85
Tigger6b	20.95	310	2.4804E-09	84

Age of DNA transposons in horse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger7	15.38	4804	2.4804E-09	62
Tigger8	36.01	802	2.4804E-09	145
Tigger9a	30.22	947	2.4804E-09	122
Tigger9b	23.78	656	2.4804E-09	96
Zaphod	27.36	1834	2.4804E-09	110
Zaphod2	28.22	751	2.4804E-09	114

Age of DNA transposons in mouse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	37.14	518	3.54E-09	105
Arthur1b	37.19	487	3.54E-09	105
BLACKJACK	35.96	468	3.54E-09	102
Charlie1	35.35	444	3.54E-09	100
Charlie10	35.43	212	3.54E-09	100
Charlie12	33.76	528	3.54E-09	95
Charlie13a	35.32	112	3.54E-09	100
Charlie18a	32.24	128	3.54E-09	91
Charlie19a	34.26	147	3.54E-09	97
Charlie1a	34.77	1336	3.54E-09	98
Charlie1b	33.94	561	3.54E-09	96
Charlie1b_Mars	33.12	1304	3.54E-09	94
Charlie2	32.86	248	3.54E-09	93
Charlie21a	32.67	103	3.54E-09	92
Charlie24	36.83	209	3.54E-09	104
Charlie25	39.37	149	3.54E-09	111
Charlie2a	35.03	251	3.54E-09	99
Charlie2b	33.29	391	3.54E-09	94
Charlie4	34.34	103	3.54E-09	97
Charlie4a	34.22	775	3.54E-09	97
Charlie4z	32.84	212	3.54E-09	93
Charlie5	31.89	958	3.54E-09	90
Charlie7	35.89	436	3.54E-09	101
Charlie8	35.18	385	3.54E-09	99
Charlie9	33.25	148	3.54E-09	94
Cheshire	35.73	295	3.54E-09	101
Cheshire_Mars	33.73	1158	3.54E-09	95
FordPrefect	39.38	178	3.54E-09	111
Kanga1	36.33	190	3.54E-09	103
Kanga11a	36.54	106	3.54E-09	103
Kanga1a	36.36	100	3.54E-09	103
Kanga1c	36.39	136	3.54E-09	103
Kanga2_a	35.77	214	3.54E-09	101
Looper	33.95	143	3.54E-09	96
MADE2	27.18	132	3.54E-09	77
MamRep1161	32.90	152	3.54E-09	93
MamRep137	36.60	148	3.54E-09	103
MamRep434	32.75	317	3.54E-09	92
MARNA	36.82	300	3.54E-09	104
MER102a	34.93	230	3.54E-09	99
MER102b	36.56	267	3.54E-09	103
MER102c	36.79	196	3.54E-09	104
MER103	31.13	433	3.54E-09	88
MER104	32.84	178	3.54E-09	93
MER105	33.56	113	3.54E-09	95
MER106B	32.80	122	3.54E-09	93
MER112	31.58	187	3.54E-09	89
MER113	33.56	341	3.54E-09	95

Age of DNA transposons in mouse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER115	39.83	303	3.54E-09	112
MER117	33.77	548	3.54E-09	95
MER119	37.34	277	3.54E-09	105
MER2	34.85	6456	3.54E-09	98
MER20	35.38	4182	3.54E-09	100
MER20B	37.05	478	3.54E-09	105
MER2B	36.29	972	3.54E-09	102
MER3	33.96	2036	3.54E-09	96
MER30	31.69	162	3.54E-09	90
MER33	31.61	1848	3.54E-09	89
MER44A	33.45	383	3.54E-09	94
MER44B	37.24	967	3.54E-09	105
MER44D	36.68	664	3.54E-09	104
MER45A	37.03	473	3.54E-09	105
MER45B	37.25	443	3.54E-09	105
MER45C	39.42	168	3.54E-09	111
MER45R	35.06	286	3.54E-09	99
MER46C	34.80	355	3.54E-09	98
MER47B	39.40	208	3.54E-09	111
MER53	29.50	1291	3.54E-09	83
MER58A	37.58	2393	3.54E-09	106
MER58B	35.14	1827	3.54E-09	99
MER58C	33.48	312	3.54E-09	95
MER58D	30.93	204	3.54E-09	87
MER5A	34.82	3264	3.54E-09	98
MER5A1	32.66	2696	3.54E-09	92
MER5B	34.42	1608	3.54E-09	97
MER5C	31.33	149	3.54E-09	88
MER63A	40.00	330	3.54E-09	113
MER63B	35.58	440	3.54E-09	100
MER63C	38.39	204	3.54E-09	108
MER63D	34.92	898	3.54E-09	99
MER81	35.78	291	3.54E-09	101
MER82	36.92	1267	3.54E-09	104
MER94	34.43	240	3.54E-09	97
MER96B	30.99	253	3.54E-09	88
MER97c	35.78	261	3.54E-09	101
MER99	35.11	126	3.54E-09	99
MMAR1	29.84	311	3.54E-09	84
ORSL	30.43	161	3.54E-09	86
ORSL-2b	32.73	148	3.54E-09	92
RCHARR1	29.63	6411	3.54E-09	84
Ricksha	36.06	641	3.54E-09	102
Ricksha_b	35.73	100	3.54E-09	101
Ricksha_c	28.85	13444	3.54E-09	81
RMER30	32.13	2851	3.54E-09	91
SPIN_NA_10_Rode	17.80	34828	3.54E-09	50
Tigger1	33.96	780	3.54E-09	96

Age of DNA transposons in mouse

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger10	37.70	101	3.54E-09	106
Tigger13a	35.99	343	3.54E-09	102
Tigger14a	33.55	124	3.54E-09	95
Tigger15a	35.31	350	3.54E-09	100
Tigger1a_Mars	33.59	109	3.54E-09	95
Tigger2b	34.38	110	3.54E-09	97
Tigger3(Golem)	35.08	3141	3.54E-09	99
Tigger3b	36.00	557	3.54E-09	102
Tigger4(Zombi)	30.67	4705	3.54E-09	87
Tigger5	38.58	1800	3.54E-09	109
Tigger5a	33.54	1379	3.54E-09	95
Tigger6a	36.49	268	3.54E-09	103
Tigger6b	34.05	209	3.54E-09	96
Tigger7	35.61	5020	3.54E-09	101
Tigger9a	37.72	104	3.54E-09	107
Tigger9b	37.70	227	3.54E-09	106
Zaphod	35.86	449	3.54E-09	101
Zaphod2	31.95	104	3.54E-09	90

Age of DNA transposons in pika

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	35.07	577	3.4214E-09	103
Arthur1A	34.28	135	3.4214E-09	100
Arthur1B	36.26	738	3.4214E-09	106
Arthur1C	33.52	110	3.4214E-09	98
BLACKJACK	35.08	580	3.4214E-09	103
Charlie1	33.26	465	3.4214E-09	97
Charlie10	34.41	325	3.4214E-09	101
Charlie13a	36.42	137	3.4214E-09	106
Charlie15a	35.11	166	3.4214E-09	103
Charlie16a	33.99	155	3.4214E-09	99
Charlie18a	32.46	261	3.4214E-09	95
Charlie19a	33.71	196	3.4214E-09	99
Charlie1a	32.72	1608	3.4214E-09	96
Charlie1b	31.84	757	3.4214E-09	93
Charlie1b_Mars	31.20	1417	3.4214E-09	91
Charlie2	32.94	334	3.4214E-09	96
Charlie20a	36.31	135	3.4214E-09	106
Charlie21a	33.19	121	3.4214E-09	97
Charlie22a	35.28	104	3.4214E-09	103
Charlie23a	33.99	102	3.4214E-09	99
Charlie24	36.08	286	3.4214E-09	105
Charlie25	38.47	220	3.4214E-09	112
Charlie2a	33.91	328	3.4214E-09	99
Charlie2b	33.06	551	3.4214E-09	97
Charlie4	33.27	107	3.4214E-09	97
Charlie4a	33.05	935	3.4214E-09	97
Charlie4z	33.74	367	3.4214E-09	99
Charlie5	30.02	1032	3.4214E-09	88
Charlie6	34.46	133	3.4214E-09	101
Charlie7	35.13	620	3.4214E-09	103
Charlie8	35.27	602	3.4214E-09	103
Charlie9	31.00	218	3.4214E-09	91
Cheshire	33.59	272	3.4214E-09	98
Cheshire_Mars	32.18	1121	3.4214E-09	94
FordPrefect	37.66	243	3.4214E-09	110
HSMAR2	27.66	7998	3.4214E-09	81
HSMAR2_NA_2_Op	23.02	46014	3.4214E-09	67
Kanga1	35.72	214	3.4214E-09	104
Kanga11a	36.66	154	3.4214E-09	107
Kanga1a	37.25	165	3.4214E-09	109
Kanga1c	34.76	221	3.4214E-09	102
Kanga1d	31.05	124	3.4214E-09	91
Kanga2_a	35.28	266	3.4214E-09	103
Looper	31.03	168	3.4214E-09	91
MADE2	26.04	238	3.4214E-09	76
MamRep1161	34.31	212	3.4214E-09	100
MamRep137	37.70	237	3.4214E-09	110
MamRep1879	36.94	116	3.4214E-09	108

Age of DNA transposons in pika

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MamRep38	39.95	166	3.4214E-09	117
MamRep4096	34.20	148	3.4214E-09	100
MamRep434	32.34	373	3.4214E-09	95
MARNA	36.99	434	3.4214E-09	108
MER102a	34.88	416	3.4214E-09	102
MER102b	37.97	497	3.4214E-09	111
MER102c	36.12	509	3.4214E-09	106
MER103	31.64	756	3.4214E-09	92
MER104	33.33	301	3.4214E-09	97
MER105	31.42	177	3.4214E-09	92
MER106A	33.31	108	3.4214E-09	97
MER106B	31.40	160	3.4214E-09	92
MER112	31.90	350	3.4214E-09	93
MER113	33.14	553	3.4214E-09	97
MER113A	32.22	169	3.4214E-09	94
MER115	40.59	387	3.4214E-09	119
MER117	34.30	661	3.4214E-09	100
MER119	35.29	363	3.4214E-09	103
MER2	32.55	7352	3.4214E-09	95
MER20	33.92	4803	3.4214E-09	99
MER20B	36.90	729	3.4214E-09	108
MER2B	34.71	1325	3.4214E-09	101
MER3	32.77	2798	3.4214E-09	96
MER33	30.33	2455	3.4214E-09	89
MER44A	32.26	617	3.4214E-09	94
MER44B	34.55	955	3.4214E-09	101
MER44C	33.72	145	3.4214E-09	99
MER44D	34.31	723	3.4214E-09	100
MER45A	36.79	729	3.4214E-09	108
MER45B	35.27	535	3.4214E-09	103
MER45C	39.30	284	3.4214E-09	115
MER45R	33.36	340	3.4214E-09	98
MER46C	34.95	510	3.4214E-09	102
MER47B	36.69	248	3.4214E-09	107
MER53	28.36	1639	3.4214E-09	83
MER58A	36.84	2959	3.4214E-09	108
MER58B	34.44	2246	3.4214E-09	101
MER58C	33.04	424	3.4214E-09	97
MER58D	29.78	280	3.4214E-09	87
MER5A	33.72	4875	3.4214E-09	99
MER5A1	32.32	3717	3.4214E-09	94
MER5B	33.82	2543	3.4214E-09	99
MER5C	31.66	238	3.4214E-09	93
MER63A	37.57	484	3.4214E-09	110
MER63B	34.73	576	3.4214E-09	102
MER63C	34.52	226	3.4214E-09	101
MER63D	32.80	978	3.4214E-09	96
MER81	35.37	411	3.4214E-09	103

Age of DNA transposons in pika

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER82	35.26	1364	3.4214E-09	103
MER91A	38.81	159	3.4214E-09	113
MER91B	33.07	145	3.4214E-09	97
MER91C	29.60	102	3.4214E-09	87
MER94	33.89	389	3.4214E-09	99
MER96	33.86	156	3.4214E-09	99
MER96B	30.03	437	3.4214E-09	88
MER97c	34.55	312	3.4214E-09	101
MER99	33.88	104	3.4214E-09	99
ORSL	30.57	223	3.4214E-09	89
ORSL-2b	31.20	143	3.4214E-09	91
Ricksha	32.95	374	3.4214E-09	96
Ricksha_c	36.71	3580	3.4214E-09	107
Tigger1	29.83	775	3.4214E-09	87
Tigger10	37.09	169	3.4214E-09	108
Tigger12c	38.28	155	3.4214E-09	112
Tigger13a	35.70	549	3.4214E-09	104
Tigger14a	33.82	173	3.4214E-09	99
Tigger15a	35.54	522	3.4214E-09	104
Tigger16a	33.68	125	3.4214E-09	98
Tigger16b	35.66	123	3.4214E-09	104
Tigger1a_Mars	31.14	117	3.4214E-09	91
Tigger3(Golem)	32.45	3758	3.4214E-09	95
Tigger3b	33.35	652	3.4214E-09	97
Tigger4(Zombi)	28.23	6407	3.4214E-09	83
Tigger5	36.25	1949	3.4214E-09	106
Tigger5a	31.13	1804	3.4214E-09	91
Tigger6a	33.56	257	3.4214E-09	98
Tigger6b	31.77	249	3.4214E-09	93
Tigger7	32.75	5328	3.4214E-09	96
Tigger8	36.64	121	3.4214E-09	107
Tigger9a	37.04	174	3.4214E-09	108
Tigger9b	34.64	262	3.4214E-09	101
Zaphod	34.18	540	3.4214E-09	100
Zaphod2	32.33	108	3.4214E-09	94

Age of DNA transposons in rabbit

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	32.65	1024	3.0369E-09	107
arthur1a	33.00	336	3.0369E-09	109
arthur1b	33.62	1393	3.0369E-09	111
Arthur1C	31.67	204	3.0369E-09	104
BLACKJACK	32.85	1087	3.0369E-09	108
Charlie1	30.77	823	3.0369E-09	101
Charlie10	31.81	605	3.0369E-09	105
Charlie10a	34.09	138	3.0369E-09	112
Charlie10b	33.25	132	3.0369E-09	109
Charlie13a	35.65	261	3.0369E-09	117
Charlie13b	36.00	134	3.0369E-09	119
Charlie14a	31.63	173	3.0369E-09	104
Charlie15a	36.03	505	3.0369E-09	119
Charlie16a	34.09	356	3.0369E-09	112
Charlie17a	37.38	243	3.0369E-09	123
Charlie18a	32.04	613	3.0369E-09	105
Charlie19a	34.58	392	3.0369E-09	114
Charlie1a	29.71	2665	3.0369E-09	98
Charlie1b	28.33	1359	3.0369E-09	93
Charlie1b_Mars	27.25	1869	3.0369E-09	90
Charlie2	31.96	621	3.0369E-09	105
Charlie20a	35.60	270	3.0369E-09	117
Charlie21a	32.32	280	3.0369E-09	106
Charlie22a	33.80	265	3.0369E-09	111
Charlie23a	34.93	272	3.0369E-09	115
Charlie24	34.89	549	3.0369E-09	115
Charlie25	37.75	495	3.0369E-09	124
Charlie26a	32.76	154	3.0369E-09	108
Charlie2a	32.47	703	3.0369E-09	107
Charlie2b	31.53	1022	3.0369E-09	104
Charlie4	31.04	122	3.0369E-09	102
Charlie4a	30.58	1619	3.0369E-09	101
Charlie4z	33.03	1196	3.0369E-09	109
Charlie5	27.24	1574	3.0369E-09	90
Charlie6	31.48	146	3.0369E-09	104
Charlie7	33.82	1344	3.0369E-09	111
Charlie7a	30.68	285	3.0369E-09	101
Charlie8	35.09	1175	3.0369E-09	116
Charlie9	31.09	454	3.0369E-09	102
Cheshire	30.21	412	3.0369E-09	99
Cheshire_Mars	28.77	1493	3.0369E-09	95
CheshMITE	30.95	273	3.0369E-09	102
FordPrefect	36.01	333	3.0369E-09	119
FordPrefect_a	41.38	107	3.0369E-09	136
HSMAR2_NA_3_Oc	19.07	86187	3.0369E-09	63
HSMAR2_Oc	21.88	12097	3.0369E-09	72
Kanga1	33.35	448	3.0369E-09	110
Kanga11a	35.80	320	3.0369E-09	118

Age of DNA transposons in rabbit

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1a	34.29	291	3.0369E-09	113
Kanga1b	33.61	194	3.0369E-09	111
Kanga1c	33.47	395	3.0369E-09	110
Kanga1d	30.69	289	3.0369E-09	101
Kanga2_a	33.53	591	3.0369E-09	110
Looper	27.96	306	3.0369E-09	92
MADE2	25.39	638	3.0369E-09	84
MamRep1161	33.92	430	3.0369E-09	112
MamRep137	37.35	528	3.0369E-09	123
MamRep1879	33.93	363	3.0369E-09	112
MamRep1894	37.16	109	3.0369E-09	122
MamRep38	39.31	471	3.0369E-09	129
MamRep4096	34.34	315	3.0369E-09	113
MamRep434	33.35	579	3.0369E-09	110
MARNA	36.89	901	3.0369E-09	121
MER102a	35.62	930	3.0369E-09	117
MER102b	37.44	1056	3.0369E-09	123
MER102c	37.03	1033	3.0369E-09	122
MER103	31.32	1712	3.0369E-09	103
MER104	31.93	646	3.0369E-09	105
MER105	29.11	309	3.0369E-09	96
MER106A	32.00	238	3.0369E-09	105
MER106B	29.05	378	3.0369E-09	96
MER112	31.01	997	3.0369E-09	102
MER113	32.28	1018	3.0369E-09	106
MER113A	31.54	373	3.0369E-09	104
MER115	39.86	845	3.0369E-09	131
MER117	35.08	1342	3.0369E-09	116
MER119	32.88	570	3.0369E-09	108
MER2	28.14	10375	3.0369E-09	93
MER20	30.22	8005	3.0369E-09	100
MER20B	35.28	1414	3.0369E-09	116
MER2B	30.83	1962	3.0369E-09	102
MER3	29.76	5091	3.0369E-09	98
MER33	27.04	4525	3.0369E-09	89
MER44A	27.57	1175	3.0369E-09	91
MER44B	30.45	1088	3.0369E-09	100
MER44C	29.01	222	3.0369E-09	96
MER44D	31.14	845	3.0369E-09	103
MER45A	33.06	1385	3.0369E-09	109
MER45B	31.24	882	3.0369E-09	103
MER45C	37.80	600	3.0369E-09	124
MER45R	29.71	590	3.0369E-09	98
MER46C	33.65	1094	3.0369E-09	111
MER47B	32.81	478	3.0369E-09	108
MER53	25.39	2831	3.0369E-09	84
MER58A	33.30	5265	3.0369E-09	110
MER58B	31.16	3587	3.0369E-09	103

Age of DNA transposons in rabbit

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER58C	31.80	1050	3.0369E-09	105
MER58D	26.34	537	3.0369E-09	87
MER5A	31.24	9074	3.0369E-09	103
MER5A1	29.95	6935	3.0369E-09	99
MER5B	32.72	5599	3.0369E-09	108
MER5C	30.21	518	3.0369E-09	99
MER5C1	32.70	227	3.0369E-09	108
MER63A	35.49	1052	3.0369E-09	117
MER63B	31.79	956	3.0369E-09	105
MER63C	31.68	308	3.0369E-09	104
MER63D	29.93	1483	3.0369E-09	99
MER81	33.05	1033	3.0369E-09	109
MER82	31.04	2009	3.0369E-09	102
MER91A	39.03	477	3.0369E-09	129
MER91B	32.73	313	3.0369E-09	108
MER91C	30.66	237	3.0369E-09	101
MER94	33.50	1063	3.0369E-09	110
MER96	32.33	353	3.0369E-09	106
MER96B	27.19	854	3.0369E-09	90
MER97c	32.84	652	3.0369E-09	108
MER99	31.23	163	3.0369E-09	103
Merlin1_HS	23.29	100	3.0369E-09	77
ORSL	27.87	430	3.0369E-09	92
ORSL-2b	31.24	267	3.0369E-09	103
Ricksha	29.66	614	3.0369E-09	98
Ricksha_b	35.99	106	3.0369E-09	119
Ricksha_c	34.65	7439	3.0369E-09	114
Tigger1	24.07	1074	3.0369E-09	79
Tigger10	36.95	336	3.0369E-09	122
Tigger11a	36.53	123	3.0369E-09	120
Tigger12	36.18	121	3.0369E-09	119
Tigger12A	35.03	148	3.0369E-09	115
Tigger12c	38.46	334	3.0369E-09	127
Tigger13a	34.97	1138	3.0369E-09	115
Tigger14a	33.47	330	3.0369E-09	110
Tigger15a	36.24	1187	3.0369E-09	119
Tigger16a	33.94	235	3.0369E-09	112
Tigger16b	34.76	254	3.0369E-09	114
Tigger1a_Art	25.97	111	3.0369E-09	86
Tigger1a_Mars	26.04	158	3.0369E-09	86
Tigger2a_Art	24.78	134	3.0369E-09	82
Tigger2b	26.20	116	3.0369E-09	86
Tigger3(Golem)	28.24	6485	3.0369E-09	93
Tigger3b	27.61	951	3.0369E-09	91
Tigger4(Zombi)	23.73	9791	3.0369E-09	78
Tigger5	32.04	3271	3.0369E-09	105
Tigger5a	27.43	3545	3.0369E-09	90
Tigger6a	29.63	386	3.0369E-09	98

Age of DNA transposons in rabbit

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger6b	27.77	362	3.0369E-09	91
Tigger7	27.91	6959	3.0369E-09	92
Tigger8	36.58	245	3.0369E-09	120
Tigger9a	34.34	337	3.0369E-09	113
Tigger9b	31.33	473	3.0369E-09	103
Zaphod	32.73	973	3.0369E-09	108
Zaphod2	31.16	223	3.0369E-09	103

Ages of DNA transposons in rat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	36.86	443	3.54E-09	104
Arthur1b	36.87	475	3.54E-09	104
BLACKJACK	36.30	416	3.54E-09	103
Charlie1	35.09	443	3.54E-09	99
Charlie10	35.67	189	3.54E-09	101
Charlie12	34.61	492	3.54E-09	98
Charlie13a	36.34	108	3.54E-09	103
Charlie18a	33.66	101	3.54E-09	95
Charlie19a	34.42	115	3.54E-09	97
Charlie1a	34.96	1235	3.54E-09	99
Charlie1b	33.74	528	3.54E-09	95
Charlie1b_Mars	33.35	1312	3.54E-09	94
Charlie2	34.00	238	3.54E-09	96
Charlie24	37.00	176	3.54E-09	104
Charlie25	39.17	154	3.54E-09	111
Charlie2a	34.96	216	3.54E-09	99
Charlie2b	33.85	361	3.54E-09	96
Charlie4a	34.69	694	3.54E-09	98
Charlie4z	33.53	184	3.54E-09	95
Charlie5	31.98	864	3.54E-09	90
Charlie7	35.84	403	3.54E-09	101
Charlie8	35.19	334	3.54E-09	99
Charlie9	32.88	113	3.54E-09	93
Cheshire	36.05	304	3.54E-09	102
Cheshire_Mars	33.56	1017	3.54E-09	95
FordPrefect	39.62	173	3.54E-09	112
Kanga1	36.92	149	3.54E-09	104
Kanga11a	36.96	111	3.54E-09	104
Kanga1c	37.55	120	3.54E-09	106
Kanga2_a	35.52	191	3.54E-09	100
Looper	33.25	113	3.54E-09	94
MADE2	27.23	114	3.54E-09	77
MamRep1161	32.26	116	3.54E-09	91
MamRep137	36.48	129	3.54E-09	103
MamRep434	33.11	291	3.54E-09	93
MARNA	37.11	277	3.54E-09	105
MER102a	35.50	208	3.54E-09	100
MER102b	36.43	248	3.54E-09	103
MER102c	37.44	196	3.54E-09	106
MER103	31.16	378	3.54E-09	88
MER104	32.84	142	3.54E-09	93
MER112	31.08	169	3.54E-09	88
MER113	33.19	326	3.54E-09	94
MER115	38.77	257	3.54E-09	109
MER117	34.57	544	3.54E-09	98
MER119	37.37	255	3.54E-09	106
MER2	35.45	5865	3.54E-09	100
MER20	35.73	3810	3.54E-09	101

Ages of DNA transposons in rat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER20B	36.75	394	3.54E-09	104
MER2B	36.54	985	3.54E-09	103
MER3	34.48	1779	3.54E-09	97
MER30	32.20	147	3.54E-09	91
MER33	31.87	1576	3.54E-09	90
MER44A	33.73	354	3.54E-09	95
MER44B	37.18	874	3.54E-09	105
MER44D	36.50	607	3.54E-09	103
MER45A	37.31	430	3.54E-09	105
MER45B	37.25	423	3.54E-09	105
MER45C	39.40	171	3.54E-09	111
MER45R	34.52	255	3.54E-09	97
MER46C	35.23	295	3.54E-09	99
MER47B	39.07	179	3.54E-09	110
MER53	29.58	1095	3.54E-09	84
MER58A	37.73	2173	3.54E-09	107
MER58B	35.43	1663	3.54E-09	100
MER58C	33.25	268	3.54E-09	94
MER58D	31.15	204	3.54E-09	88
MER5A	34.94	2971	3.54E-09	99
MER5A1	32.63	2332	3.54E-09	92
MER5B	34.79	1442	3.54E-09	98
MER5C	31.25	130	3.54E-09	88
MER63A	40.04	277	3.54E-09	113
MER63B	36.15	384	3.54E-09	102
MER63C	38.32	183	3.54E-09	108
MER63D	35.20	764	3.54E-09	99
MER81	36.46	269	3.54E-09	103
MER82	37.33	1162	3.54E-09	105
MER94	33.80	232	3.54E-09	95
MER96B	30.71	243	3.54E-09	87
MER97c	36.07	222	3.54E-09	102
MER99	35.68	132	3.54E-09	101
MMAR1	30.57	331	3.54E-09	86
ORSL	30.25	160	3.54E-09	85
ORSL-2b	32.63	123	3.54E-09	92
RCHARR1	30.10	6215	3.54E-09	85
Ricksha	36.12	597	3.54E-09	102
Ricksha_c	33.60	7006	3.54E-09	95
RMER30	32.55	2619	3.54E-09	92
SPIN_NA_10_Rode	18.37	33376	3.54E-09	52
Tigger1	34.16	776	3.54E-09	96
Tigger10	37.69	103	3.54E-09	106
Tigger13a	36.16	307	3.54E-09	102
Tigger14a	34.11	105	3.54E-09	96
Tigger15a	35.65	288	3.54E-09	101
Tigger1a_Mars	33.24	125	3.54E-09	94
Tigger2b	35.53	129	3.54E-09	100

Ages of DNA transposons in rat

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger3(Golem)	35.48	2731	3.54E-09	100
Tigger3b	36.40	517	3.54E-09	103
Tigger4(Zombi)	30.99	4372	3.54E-09	88
Tigger5	38.65	1625	3.54E-09	109
Tigger5a	34.24	1270	3.54E-09	97
Tigger6a	35.87	222	3.54E-09	101
Tigger6b	35.59	206	3.54E-09	101
Tigger7	35.95	4631	3.54E-09	102
Tigger9b	37.91	206	3.54E-09	107
Zaphod	36.19	409	3.54E-09	102

Age of DNA transposons in squirrel

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	31.56	1088	2.9928E-09	105
Arthur1A	31.97	444	2.9928E-09	107
Arthur1B	33.43	1638	2.9928E-09	112
Arthur1C	30.12	247	2.9928E-09	101
BLACKJACK	32.24	1252	2.9928E-09	108
Charlie1	29.71	776	2.9928E-09	99
Charlie10	31.40	772	2.9928E-09	105
Charlie10a	33.79	152	2.9928E-09	113
Charlie10b	32.03	131	2.9928E-09	107
Charlie13a	34.83	270	2.9928E-09	116
Charlie13b	35.21	165	2.9928E-09	118
Charlie14a	30.83	212	2.9928E-09	103
Charlie15a	35.82	676	2.9928E-09	120
Charlie16a	33.48	483	2.9928E-09	112
Charlie17a	36.73	368	2.9928E-09	123
Charlie18a	31.04	775	2.9928E-09	104
Charlie19a	33.43	490	2.9928E-09	112
Charlie1a	28.41	2709	2.9928E-09	95
Charlie1b	27.63	1505	2.9928E-09	92
Charlie1b_Mars	26.63	1952	2.9928E-09	89
Charlie2	31.27	748	2.9928E-09	104
Charlie20a	34.65	344	2.9928E-09	116
Charlie21a	32.28	305	2.9928E-09	108
Charlie22a	33.26	324	2.9928E-09	111
Charlie23a	33.96	385	2.9928E-09	113
Charlie24	34.58	635	2.9928E-09	116
Charlie25	36.55	532	2.9928E-09	122
Charlie26a	31.13	191	2.9928E-09	104
Charlie2a	31.16	729	2.9928E-09	104
Charlie2b	30.58	1235	2.9928E-09	102
Charlie4	29.67	120	2.9928E-09	99
Charlie4a	29.14	1818	2.9928E-09	97
Charlie4z	32.56	1636	2.9928E-09	109
Charlie5	25.88	1637	2.9928E-09	86
Charlie6	30.25	327	2.9928E-09	101
Charlie7	33.10	1574	2.9928E-09	111
Charlie7a	29.62	418	2.9928E-09	99
Charlie8	34.34	1384	2.9928E-09	115
Charlie9	29.05	509	2.9928E-09	97
Cheshire	28.61	356	2.9928E-09	96
Cheshire_Mars	28.12	1434	2.9928E-09	94
CheshMITE	30.52	275	2.9928E-09	102
FordPrefect	35.27	343	2.9928E-09	118
FordPrefect_a	41.73	147	2.9928E-09	139
Kanga1	32.14	522	2.9928E-09	107
Kanga11a	35.16	332	2.9928E-09	117
Kanga1a	33.06	335	2.9928E-09	110
Kanga1b	33.43	202	2.9928E-09	112

Age of DNA transposons in squirrel

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1c	32.58	481	2.9928E-09	109
Kanga1d	30.14	274	2.9928E-09	101
Kanga2_a	32.33	589	2.9928E-09	108
Looper	26.80	282	2.9928E-09	90
MADE2	24.39	833	2.9928E-09	81
MamRep1161	33.30	511	2.9928E-09	111
MamRep137	36.99	722	2.9928E-09	124
MamRep1879	33.89	467	2.9928E-09	113
MamRep1894	37.86	124	2.9928E-09	126
MamRep38	38.93	571	2.9928E-09	130
MamRep4096	32.48	429	2.9928E-09	109
MamRep434	33.01	727	2.9928E-09	110
MARNA	36.52	1109	2.9928E-09	122
MER102a	34.81	1113	2.9928E-09	116
MER102b	37.05	1286	2.9928E-09	124
MER102c	37.08	1256	2.9928E-09	124
MER103	30.38	2168	2.9928E-09	101
MER104	30.55	881	2.9928E-09	102
MER105	27.90	327	2.9928E-09	93
MER106A	31.11	291	2.9928E-09	104
MER106B	27.73	430	2.9928E-09	93
MER112	29.90	1276	2.9928E-09	100
MER113	31.47	1236	2.9928E-09	105
MER113A	30.58	533	2.9928E-09	102
MER115	39.38	971	2.9928E-09	132
MER117	34.81	1407	2.9928E-09	116
MER119	31.88	651	2.9928E-09	107
MER2	26.59	11167	2.9928E-09	89
MER20	29.17	9483	2.9928E-09	97
MER20B	34.14	1858	2.9928E-09	114
MER2B	29.94	2023	2.9928E-09	100
MER3	28.25	6252	2.9928E-09	94
MER33	25.69	5209	2.9928E-09	86
MER44A	26.50	1389	2.9928E-09	89
MER44B	28.17	1274	2.9928E-09	94
MER44C	28.35	269	2.9928E-09	95
MER44D	30.04	953	2.9928E-09	100
MER45A	32.08	1629	2.9928E-09	107
MER45B	30.68	952	2.9928E-09	103
MER45C	36.74	664	2.9928E-09	123
MER45R	29.26	626	2.9928E-09	98
MER46C	32.34	1265	2.9928E-09	108
MER47B	31.22	770	2.9928E-09	104
MER53	23.80	3395	2.9928E-09	80
MER58A	32.50	5947	2.9928E-09	109
MER58B	30.04	4067	2.9928E-09	100
MER58C	30.96	1131	2.9928E-09	103
MER58D	25.77	666	2.9928E-09	86

Age of DNA transposons in squirrel

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER5A	30.17	11062	2.9928E-09	101
MER5A1	28.60	8604	2.9928E-09	96
MER5B	31.93	7336	2.9928E-09	107
MER5C	29.52	662	2.9928E-09	99
MER5C1	32.64	254	2.9928E-09	109
MER63A	34.51	1248	2.9928E-09	115
MER63B	30.12	1240	2.9928E-09	101
MER63C	29.71	333	2.9928E-09	99
MER63D	29.01	1679	2.9928E-09	97
MER81	32.44	1331	2.9928E-09	108
MER82	29.92	2178	2.9928E-09	100
MER91A	38.74	545	2.9928E-09	129
MER91B	32.56	381	2.9928E-09	109
MER91C	31.13	203	2.9928E-09	104
MER94	32.50	1385	2.9928E-09	109
MER96	31.79	396	2.9928E-09	106
MER96B	26.23	1076	2.9928E-09	88
MER97c	31.58	790	2.9928E-09	106
MER99	29.61	180	2.9928E-09	99
ORSL	26.62	636	2.9928E-09	89
ORSL-2a	31.68	133	2.9928E-09	106
ORSL-2b	29.67	331	2.9928E-09	99
Ricksha	28.61	759	2.9928E-09	96
Ricksha_b	34.61	115	2.9928E-09	116
Ricksha_c	35.25	11427	2.9928E-09	118
Tigger1	23.61	1179	2.9928E-09	79
Tigger10	37.78	389	2.9928E-09	126
Tigger11a	35.91	149	2.9928E-09	120
Tigger12	34.79	147	2.9928E-09	116
Tigger12A	35.11	150	2.9928E-09	117
Tigger12c	37.84	422	2.9928E-09	126
Tigger13a	34.51	1321	2.9928E-09	115
Tigger14a	32.76	423	2.9928E-09	109
Tigger15a	36.18	1362	2.9928E-09	121
Tigger16a	33.89	242	2.9928E-09	113
Tigger16b	34.75	300	2.9928E-09	116
Tigger1a_Mars	25.12	183	2.9928E-09	84
Tigger2a_Art	24.63	170	2.9928E-09	82
Tigger2b	26.20	260	2.9928E-09	88
Tigger3(Golem)	26.86	7202	2.9928E-09	90
Tigger3b	26.36	965	2.9928E-09	88
Tigger4(Zombi)	22.66	10286	2.9928E-09	76
Tigger5	30.43	3543	2.9928E-09	102
Tigger5a	26.59	4047	2.9928E-09	89
Tigger6a	28.71	394	2.9928E-09	96
Tigger6b	27.05	328	2.9928E-09	90
Tigger7	26.54	7295	2.9928E-09	89
Tigger8	35.99	292	2.9928E-09	120

Age of DNA transposons in squirrel

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger9a	34.26	432	2.9928E-09	114
Tigger9b	30.02	477	2.9928E-09	100
Zaphod	31.58	1062	2.9928E-09	106
Zaphod2	30.18	294	2.9928E-09	101

Age of DNA transposons in tenrec

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	33.68	649	2.9173E-09	115
arthur1a	32.79	35	2.9173E-09	112
arthur1b	35.18	240	2.9173E-09	121
Arthur1C	31.88	172	2.9173E-09	109
BLACKJACK	33.01	2160	2.9173E-09	113
Charlie1	31.17	507	2.9173E-09	107
Charlie10	33.38	374	2.9173E-09	114
Charlie13a	35.56	187	2.9173E-09	122
Charlie13b	34.97	103	2.9173E-09	120
Charlie15a	35.30	263	2.9173E-09	121
Charlie16a	33.89	217	2.9173E-09	116
Charlie17a	37.37	146	2.9173E-09	128
Charlie18a	31.79	312	2.9173E-09	109
Charlie19a	35.01	245	2.9173E-09	120
Charlie1a	30.92	1907	2.9173E-09	106
Charlie1b	29.37	910	2.9173E-09	101
Charlie1b_Mars	29.00	1442	2.9173E-09	99
Charlie2	32.56	470	2.9173E-09	112
Charlie20a	34.96	183	2.9173E-09	120
Charlie21a	32.54	196	2.9173E-09	112
Charlie22a	33.74	164	2.9173E-09	116
Charlie23a	34.59	193	2.9173E-09	119
Charlie24	34.75	417	2.9173E-09	119
Charlie25	37.58	330	2.9173E-09	129
Charlie26a	31.92	106	2.9173E-09	109
Charlie2a	32.98	480	2.9173E-09	113
Charlie2b	32.09	711	2.9173E-09	110
Charlie4a	31.51	1146	2.9173E-09	108
Charlie4z	32.78	737	2.9173E-09	112
Charlie5	28.92	1111	2.9173E-09	99
Charlie6	32.45	133	2.9173E-09	111
Charlie7	34.04	908	2.9173E-09	117
Charlie7a	30.27	162	2.9173E-09	104
Charlie8	35.40	755	2.9173E-09	121
Charlie9	33.87	321	2.9173E-09	116
Cheshire	30.27	312	2.9173E-09	104
Cheshire_Mars	30.10	1111	2.9173E-09	103
CheshMITE	28.86	141	2.9173E-09	99
Dumbo	26.68	2362	2.9173E-09	91
Dumbo_NA_1	28.62	1897	2.9173E-09	98
Dumbo_NA_2	26.74	2086	2.9173E-09	92
FordPrefect	37.19	234	2.9173E-09	127
HSMAR2_Et	25.98	5503	2.9173E-09	89
HSMAR2_NA_4_Et	20.96	42136	2.9173E-09	72
Kanga1	34.30	305	2.9173E-09	118
Kanga11a	34.94	228	2.9173E-09	120
Kanga1a	34.96	219	2.9173E-09	120
Kanga1b	33.89	137	2.9173E-09	116

Age of DNA transposons in tenrec

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga1c	33.31	241	2.9173E-09	114
Kanga1d	30.59	194	2.9173E-09	105
Kanga2_a	33.81	379	2.9173E-09	116
LAMAR2_Et	26.21	27458	2.9173E-09	90
LAMAR2_NA_1_Et	26.40	26853	2.9173E-09	90
Looper	29.47	250	2.9173E-09	101
MADE2	26.28	356	2.9173E-09	90
MamRep1161	34.17	284	2.9173E-09	117
MamRep137	36.91	387	2.9173E-09	127
MamRep1879	34.57	160	2.9173E-09	118
MamRep38	38.67	252	2.9173E-09	133
MamRep4096	32.62	222	2.9173E-09	112
MamRep434	33.49	466	2.9173E-09	115
MARNA	41.49	1445	2.9173E-09	142
MER102a	35.44	618	2.9173E-09	121
MER102b	36.80	752	2.9173E-09	126
MER102c	36.61	802	2.9173E-09	125
MER103	31.46	1017	2.9173E-09	108
MER104	32.49	402	2.9173E-09	111
MER105	29.47	214	2.9173E-09	101
MER106A	31.89	144	2.9173E-09	109
MER106B	28.88	214	2.9173E-09	99
MER112	30.89	558	2.9173E-09	106
MER113	32.61	649	2.9173E-09	112
MER113A	32.01	211	2.9173E-09	110
MER115	39.17	559	2.9173E-09	134
MER117	34.98	950	2.9173E-09	120
MER119	33.25	439	2.9173E-09	114
MER2	27.60	897	2.9173E-09	95
MER20	30.25	6263	2.9173E-09	104
MER20B	35.50	1001	2.9173E-09	122
MER2B	31.09	452	2.9173E-09	107
MER3	30.33	3573	2.9173E-09	104
MER33	28.25	3030	2.9173E-09	97
MER45A	33.39	960	2.9173E-09	114
MER45B	33.36	592	2.9173E-09	114
MER45C	38.17	409	2.9173E-09	131
MER45R	31.38	392	2.9173E-09	108
MER46C	33.53	740	2.9173E-09	115
MER53	26.03	1965	2.9173E-09	89
MER58A	34.12	3568	2.9173E-09	117
MER58B	32.00	2729	2.9173E-09	110
MER58C	31.79	597	2.9173E-09	109
MER58D	27.41	376	2.9173E-09	94
MER5A	31.15	6817	2.9173E-09	107
MER5A1	28.58	7350	2.9173E-09	98
MER5B	32.65	4209	2.9173E-09	112
MER5C	31.33	331	2.9173E-09	107

Age of DNA transposons in tenrec

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER5C1	33.60	145	2.9173E-09	115
MER63A	35.86	681	2.9173E-09	123
MER63B	31.78	703	2.9173E-09	109
MER63C	32.73	189	2.9173E-09	112
MER63D	30.23	1157	2.9173E-09	104
MER81	33.34	640	2.9173E-09	114
MER82	31.43	1262	2.9173E-09	108
MER91A	37.98	222	2.9173E-09	130
MER91B	33.09	191	2.9173E-09	113
MER94	30.02	956	2.9173E-09	103
MER96	32.05	243	2.9173E-09	110
MER96B	28.84	553	2.9173E-09	99
MER97c	33.10	466	2.9173E-09	113
MER99	32.22	118	2.9173E-09	110
OposCharlie1	15.45	6505	2.9173E-09	53
OposCharlie1_NA	8.21	7328	2.9173E-09	28
ORSL	28.63	289	2.9173E-09	98
ORSL-2b	32.31	193	2.9173E-09	111
RCHARR1_Et	24.61	1223	2.9173E-09	84
Ricksha	29.40	267	2.9173E-09	101
Ricksha_c	25.99	11407	2.9173E-09	89
SPIN_Et	10.05	13960	2.9173E-09	34
SPIN_NA_1_Et	11.23	52541	2.9173E-09	39
SPIN_NA_6_Et	10.07	32817	2.9173E-09	35
Tigger1	27.81	3036	2.9173E-09	95
Tigger10	37.15	200	2.9173E-09	127
Tigger11a	37.51	310	2.9173E-09	129
Tigger12	39.87	217	2.9173E-09	137
Tigger12A	34.26	103	2.9173E-09	117
Tigger12c	37.03	231	2.9173E-09	127
Tigger13a	35.74	744	2.9173E-09	123
Tigger14a	33.52	245	2.9173E-09	115
Tigger15a	35.78	695	2.9173E-09	123
Tigger16a	33.52	177	2.9173E-09	115
Tigger16b	34.65	199	2.9173E-09	119
Tigger1a_Art	27.89	1841	2.9173E-09	96
Tigger1a_Car	27.22	1291	2.9173E-09	93
Tigger1a_Mars	26.58	2149	2.9173E-09	91
Tigger2	27.75	457	2.9173E-09	95
Tigger2a	26.48	232	2.9173E-09	91
Tigger2a_Art	26.60	194	2.9173E-09	91
Tigger2a_Car	25.01	312	2.9173E-09	86
Tigger2b	27.11	310	2.9173E-09	93
Tigger2b_Pri	29.92	416	2.9173E-09	103
Tigger2f	30.84	1527	2.9173E-09	106
Tigger3(Golem)	39.30	683	2.9173E-09	135
Tigger6a	30.43	277	2.9173E-09	104
Tigger6b	29.40	251	2.9173E-09	101

Age of DNA transposons in tenrec

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger7	31.91	423	2.9173E-09	109
Tigger8	35.70	175	2.9173E-09	122
Tigger9a	34.91	242	2.9173E-09	120
Tigger9b	32.75	342	2.9173E-09	112
Zaphod	33.38	712	2.9173E-09	114
Zaphod2	31.42	171	2.9173E-09	108

Age of DNA transposons in treeshrew

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Arthur1	32.18	1281	2.9551E-09	109
Arthur1A	32.59	393	2.9551E-09	110
Arthur1B	33.58	1289	2.9551E-09	114
Arthur1C	29.07	409	2.9551E-09	98
BLACKJACK	32.44	1271	2.9551E-09	110
Charlie1	27.62	2391	2.9551E-09	93
Charlie10	31.38	717	2.9551E-09	106
Charlie10a	33.14	141	2.9551E-09	112
Charlie10b	31.94	136	2.9551E-09	108
Charlie13a	35.85	290	2.9551E-09	121
Charlie13b	34.42	162	2.9551E-09	116
Charlie14a	30.78	211	2.9551E-09	104
Charlie15a	36.07	689	2.9551E-09	122
Charlie16a	33.51	468	2.9551E-09	113
Charlie17a	37.00	327	2.9551E-09	125
Charlie18a	31.75	732	2.9551E-09	107
Charlie1a	29.08	1796	2.9551E-09	98
Charlie1b	27.48	1221	2.9551E-09	93
Charlie1b_Mars	26.50	2000	2.9551E-09	90
Charlie2	32.17	659	2.9551E-09	109
Charlie20a	35.51	308	2.9551E-09	120
Charlie21a	31.82	296	2.9551E-09	108
Charlie22a	33.32	355	2.9551E-09	113
Charlie23a	34.30	391	2.9551E-09	116
Charlie24	34.60	657	2.9551E-09	117
Charlie25	36.76	555	2.9551E-09	124
Charlie26a	31.68	187	2.9551E-09	107
Charlie2a	30.47	547	2.9551E-09	103
Charlie2b	30.77	1563	2.9551E-09	104
Charlie3	23.40	1219	2.9551E-09	79
Charlie4	30.72	121	2.9551E-09	104
Charlie4a	29.16	1959	2.9551E-09	99
Charlie4z	33.14	1707	2.9551E-09	112
Charlie5	25.01	3447	2.9551E-09	85
Charlie6	31.88	174	2.9551E-09	108
Charlie7	33.40	1569	2.9551E-09	113
Charlie7a	29.88	365	2.9551E-09	101
Charlie8	34.77	1373	2.9551E-09	118
Charlie9	29.61	605	2.9551E-09	100
Cheshire	29.10	392	2.9551E-09	98
Cheshire_Mars	28.41	1476	2.9551E-09	96
CheshMITE	30.81	320	2.9551E-09	104
FordPrefect	35.91	414	2.9551E-09	122
FordPrefect_a	41.90	165	2.9551E-09	142
HSMAR1_Tb	18.56	1897	2.9551E-09	63
HSMAR2_NA_1_Tb	19.27	54342	2.9551E-09	65
HSMAR2_Tb	20.10	9878	2.9551E-09	68
Kanga1	32.36	526	2.9551E-09	110

Age of DNA transposons in treeshrew

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Kanga11a	34.81	364	2.9551E-09	118
Kanga1a	33.33	349	2.9551E-09	113
Kanga1b	31.55	239	2.9551E-09	107
Kanga1c	33.25	372	2.9551E-09	113
Kanga1d	29.80	334	2.9551E-09	101
Kanga2_a	32.06	740	2.9551E-09	109
Looper	27.19	337	2.9551E-09	92
MADE2	24.83	828	2.9551E-09	84
MamRep1161	33.93	542	2.9551E-09	115
MamRep137	36.84	782	2.9551E-09	125
MamRep1879	34.67	385	2.9551E-09	117
MamRep1894	36.97	145	2.9551E-09	125
MamRep38	38.97	609	2.9551E-09	132
MamRep4096	32.83	439	2.9551E-09	111
MamRep434	33.06	772	2.9551E-09	112
MARNA	36.92	1114	2.9551E-09	125
MER102a	35.32	1134	2.9551E-09	120
MER102b	37.36	1307	2.9551E-09	126
MER102c	36.90	1445	2.9551E-09	125
MER103	30.58	2327	2.9551E-09	103
MER104	29.45	936	2.9551E-09	100
MER105	28.19	403	2.9551E-09	95
MER106A	30.40	319	2.9551E-09	103
MER106B	27.60	417	2.9551E-09	93
MER112	31.74	1071	2.9551E-09	107
MER113	31.45	1359	2.9551E-09	106
MER113A	30.84	512	2.9551E-09	104
MER115	39.85	967	2.9551E-09	135
MER117	35.55	1618	2.9551E-09	120
MER119	31.47	687	2.9551E-09	106
MER1A_Tb_1	21.55	831	2.9551E-09	73
MER1A_Tb_2	25.06	792	2.9551E-09	85
MER1A_Tb_3	23.31	1544	2.9551E-09	79
MER1A_Tb_4	23.10	1536	2.9551E-09	78
MER2	25.67	3594	2.9551E-09	87
MER20	29.18	9362	2.9551E-09	99
MER20B	34.83	1754	2.9551E-09	118
MER2B	28.74	1888	2.9551E-09	97
MER3	28.70	5860	2.9551E-09	97
MER33	26.10	3246	2.9551E-09	88
MER44A	24.06	2309	2.9551E-09	81
MER44B	27.81	672	2.9551E-09	94
MER44C	26.67	227	2.9551E-09	90
MER45A	31.20	1584	2.9551E-09	106
MER45B	29.47	949	2.9551E-09	100
MER45C	36.70	662	2.9551E-09	124
MER45R	29.57	733	2.9551E-09	100
MER46C	32.60	1258	2.9551E-09	110

Age of DNA transposons in treeshrew

TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
MER47B	29.62	292	2.9551E-09	100
MER47C	27.69	232	2.9551E-09	94
MER53	23.73	3331	2.9551E-09	80
MER58A	32.40	6025	2.9551E-09	110
MER58B	29.95	4125	2.9551E-09	101
MER58C	30.58	1187	2.9551E-09	103
MER58D	25.54	649	2.9551E-09	86
MER5A	29.57	11622	2.9551E-09	100
MER5A1	28.52	8513	2.9551E-09	97
MER5B	30.97	6005	2.9551E-09	105
MER5C	28.08	589	2.9551E-09	95
MER5C1	32.03	309	2.9551E-09	108
MER63A	33.92	1403	2.9551E-09	115
MER63B	29.61	1289	2.9551E-09	100
MER63C	29.43	316	2.9551E-09	100
MER63D	28.28	1828	2.9551E-09	96
MER81	32.28	1373	2.9551E-09	109
MER82	29.17	2147	2.9551E-09	99
MER91A	38.78	505	2.9551E-09	131
MER91B	32.27	394	2.9551E-09	109
MER91C	31.70	223	2.9551E-09	107
MER94	32.45	1494	2.9551E-09	110
MER96	30.90	469	2.9551E-09	105
MER96B	26.27	1002	2.9551E-09	89
MER97c	31.50	819	2.9551E-09	107
MER99	29.23	169	2.9551E-09	99
Merlin1_HS	22.61	115	2.9551E-09	77
ORSL	25.90	524	2.9551E-09	88
ORSL-2a	31.31	194	2.9551E-09	106
ORSL-2b	30.21	315	2.9551E-09	102
Ricksha	34.43	485	2.9551E-09	117
Ricksha_0	23.51	176	2.9551E-09	80
Ricksha_a	23.82	123	2.9551E-09	81
Ricksha_c	33.89	10858	2.9551E-09	115
Tigger1	22.07	3186	2.9551E-09	75
Tigger10	37.73	344	2.9551E-09	128
Tigger11a	35.56	146	2.9551E-09	120
Tigger12	35.46	156	2.9551E-09	120
Tigger12A	34.96	184	2.9551E-09	118
Tigger12c	37.87	450	2.9551E-09	128
Tigger13a	34.65	1560	2.9551E-09	117
Tigger14a	32.73	411	2.9551E-09	111
Tigger15a	36.30	1462	2.9551E-09	123
Tigger16b	33.41	424	2.9551E-09	113
Tigger1a_Art	22.51	176	2.9551E-09	76
Tigger1a_Car	21.99	117	2.9551E-09	74
Tigger1a_Mars	24.11	299	2.9551E-09	82
Tigger2	24.41	140	2.9551E-09	83

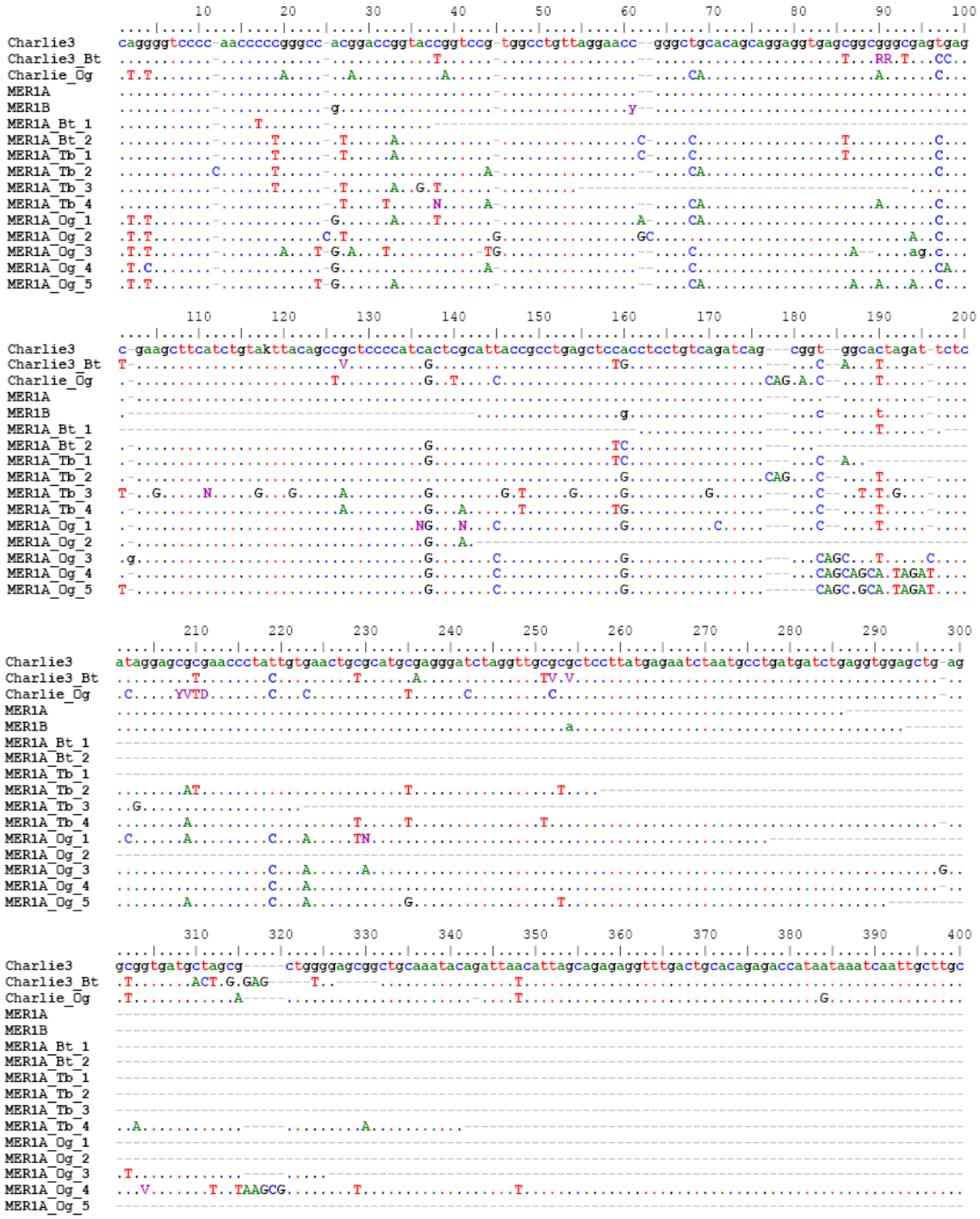
Age of DNA transposons in treeshrew

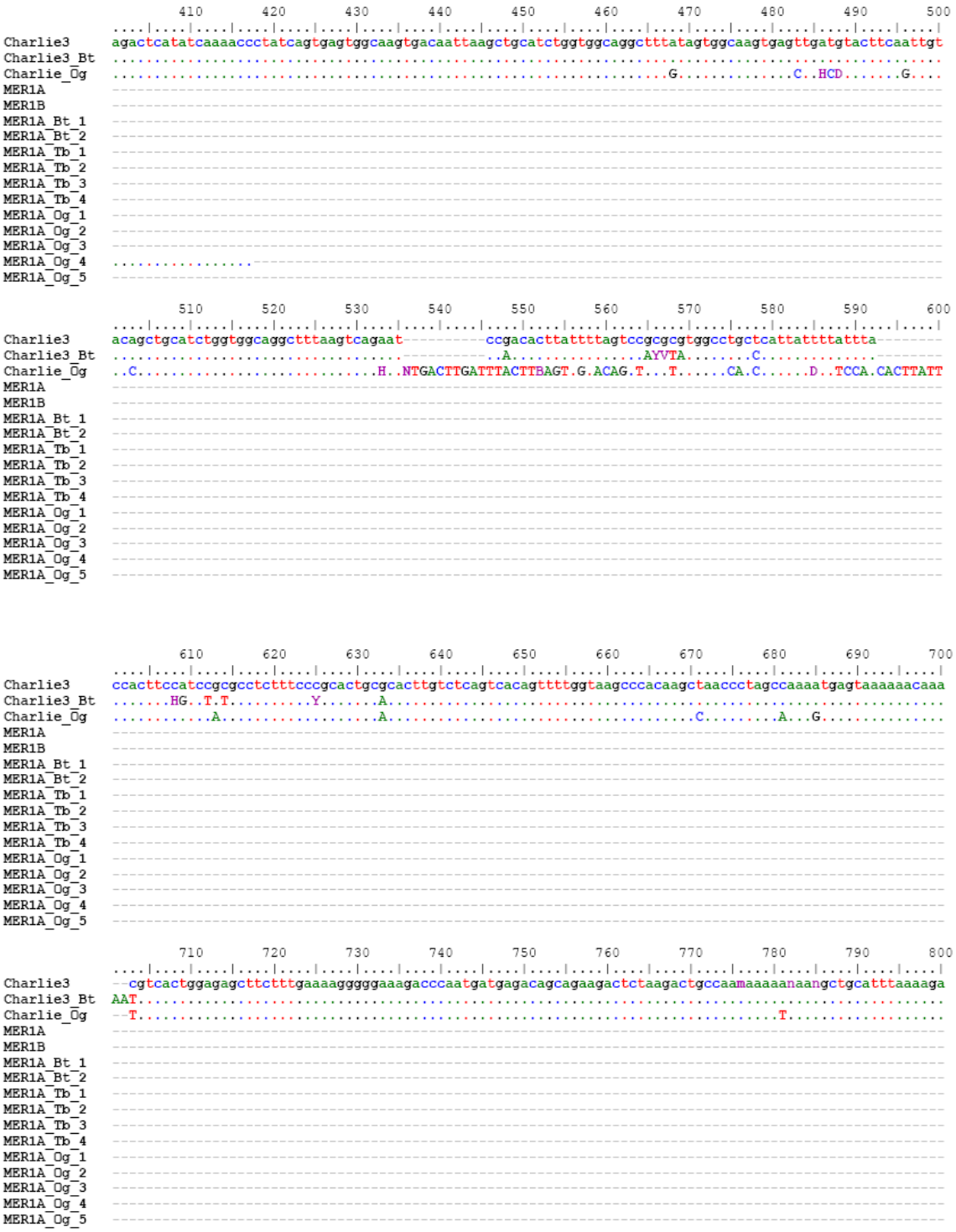
TE Name	Percent Divergence	Copy Number	Median Substitution Rate (Substitutions per year)	Age
Tigger2a_Art	23.82	133	2.9551E-09	81
Tigger2a_Car	22.59	1009	2.9551E-09	76
Tigger2b	25.81	142	2.9551E-09	87
Tigger2b_Pri	24.51	292	2.9551E-09	83
Tigger2f	22.02	683	2.9551E-09	75
Tigger3(Golem)	23.99	3326	2.9551E-09	81
Tigger3b	25.66	8740	2.9551E-09	87
Tigger4(Zombi)	22.54	11384	2.9551E-09	76
Tigger5	28.76	836	2.9551E-09	97
Tigger5a	24.03	3635	2.9551E-09	81
Tigger6a	28.26	441	2.9551E-09	96
Tigger6b	26.85	374	2.9551E-09	91
Tigger7	26.35	4480	2.9551E-09	89
Tigger8	37.09	278	2.9551E-09	125
Tigger9a	33.95	452	2.9551E-09	115
Tigger9b	29.79	489	2.9551E-09	101
Zaphod	31.60	1100	2.9551E-09	107
Zaphod2	30.89	310	2.9551E-09	105

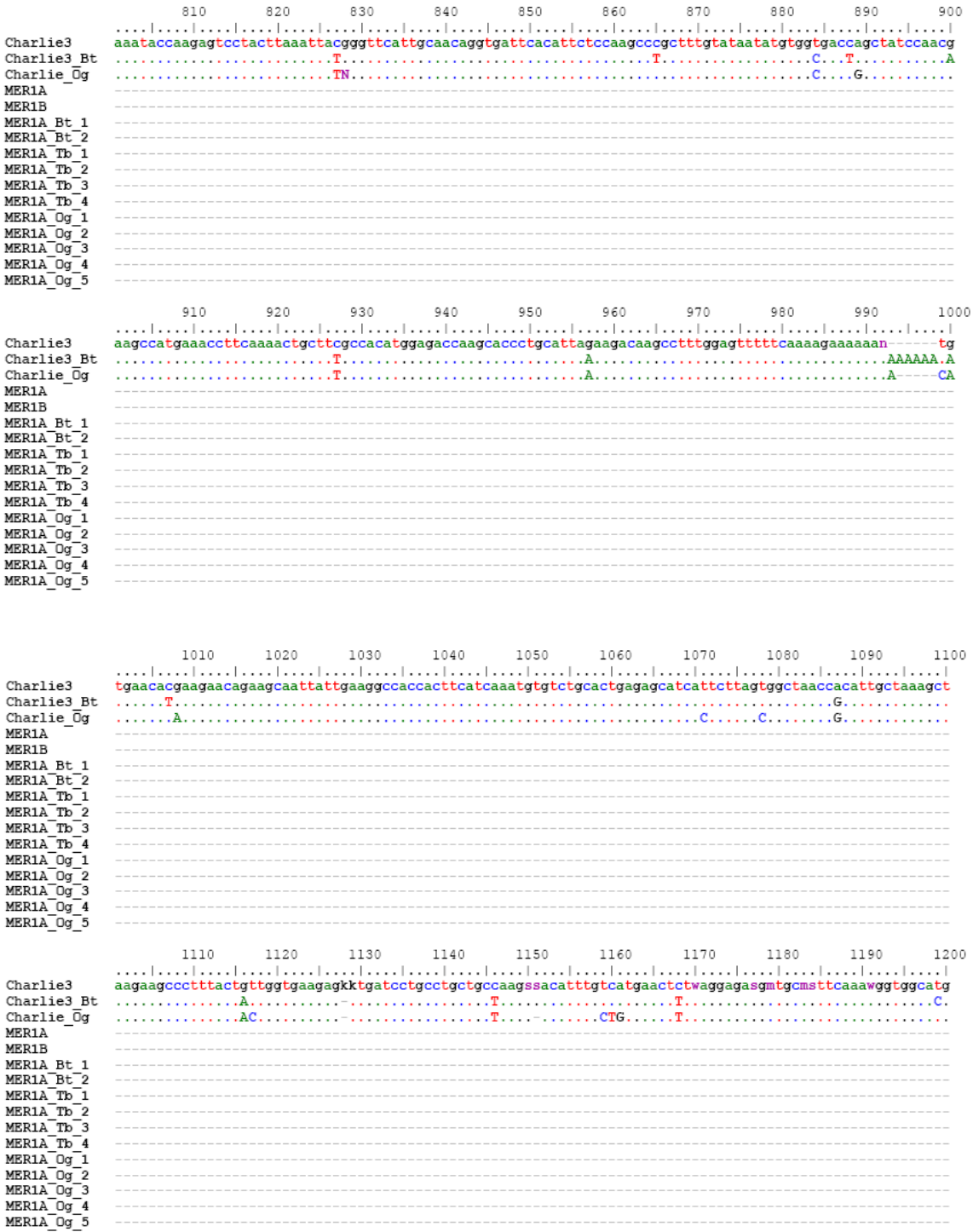
APPENDIX E

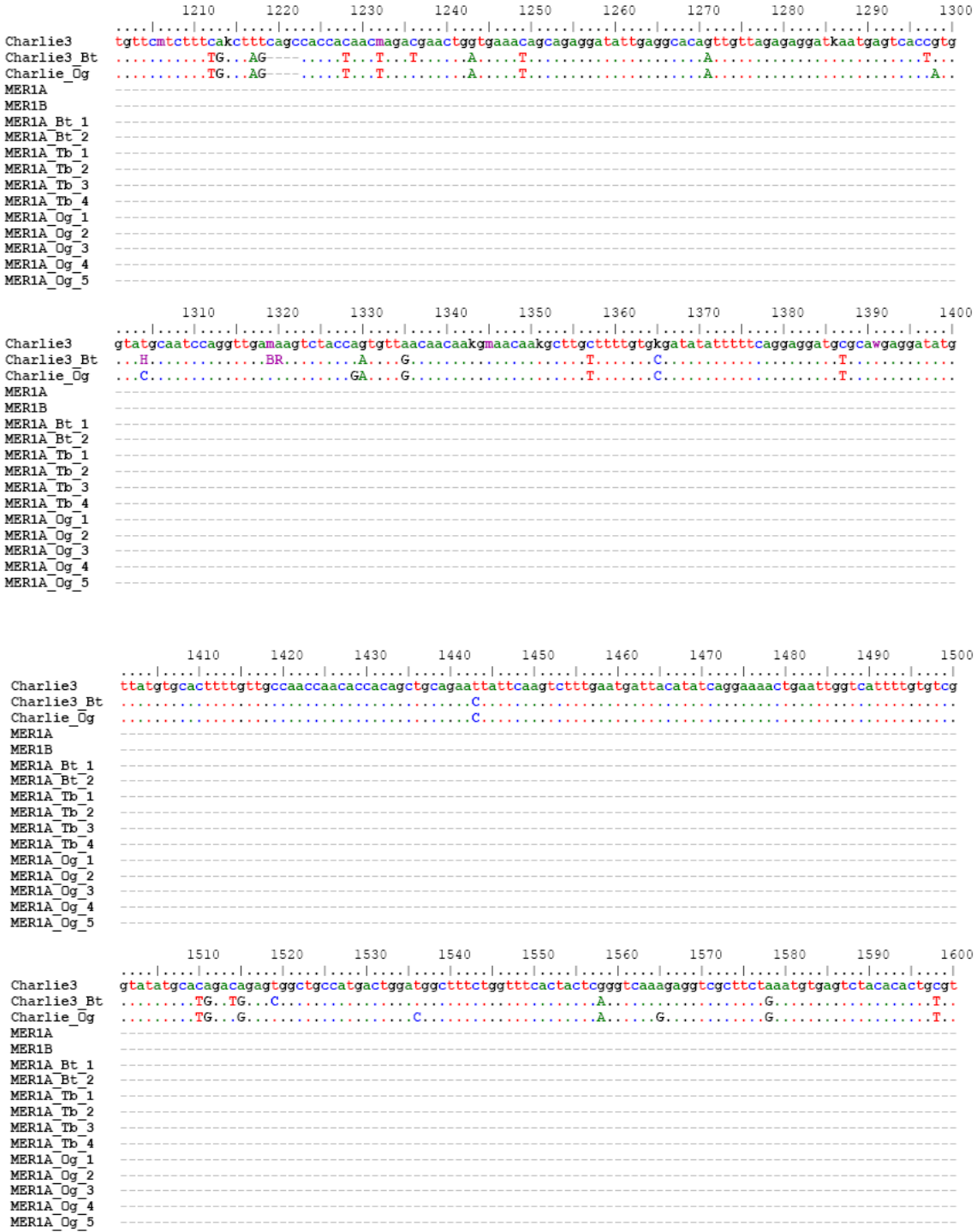
ALIGNMENTS OF HORIZONTALLY TRANSFERRED MAMMALIAN DNA TRIBES. THIS DOCUMENTS IS ALSO AVAILABLE ONLINE AT http://feschottelab.uta.edu/pace/dna_tribe_alignments.pdf. THE ONLINE VERSION INCLUDES NOT ONLY THE HORIZONTALLY TRANSFERRED TRIBES, BUT ALL OTHER MAMMALIAN DNA TE TRIBES AS WELL.

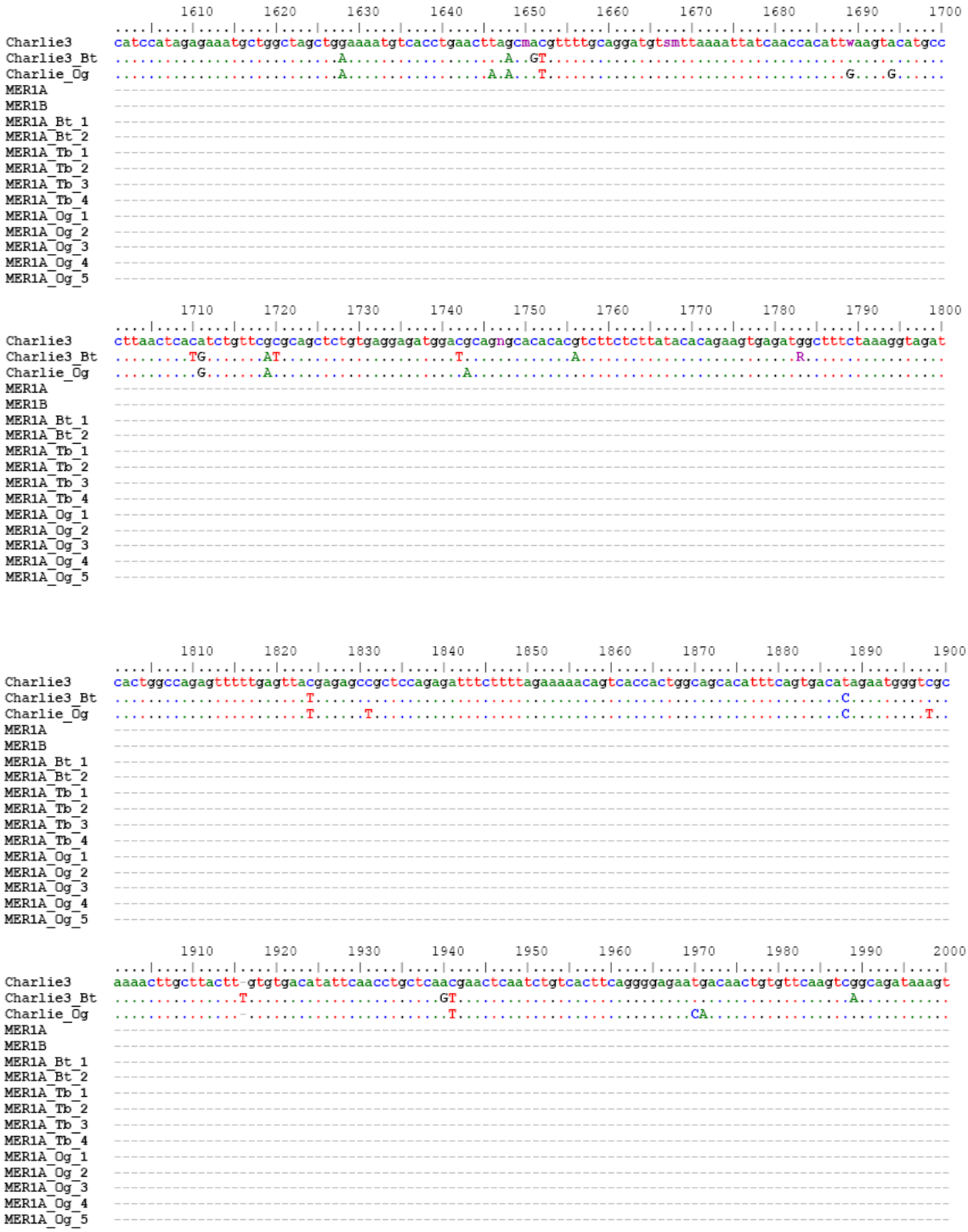
Charlie3 Tribe

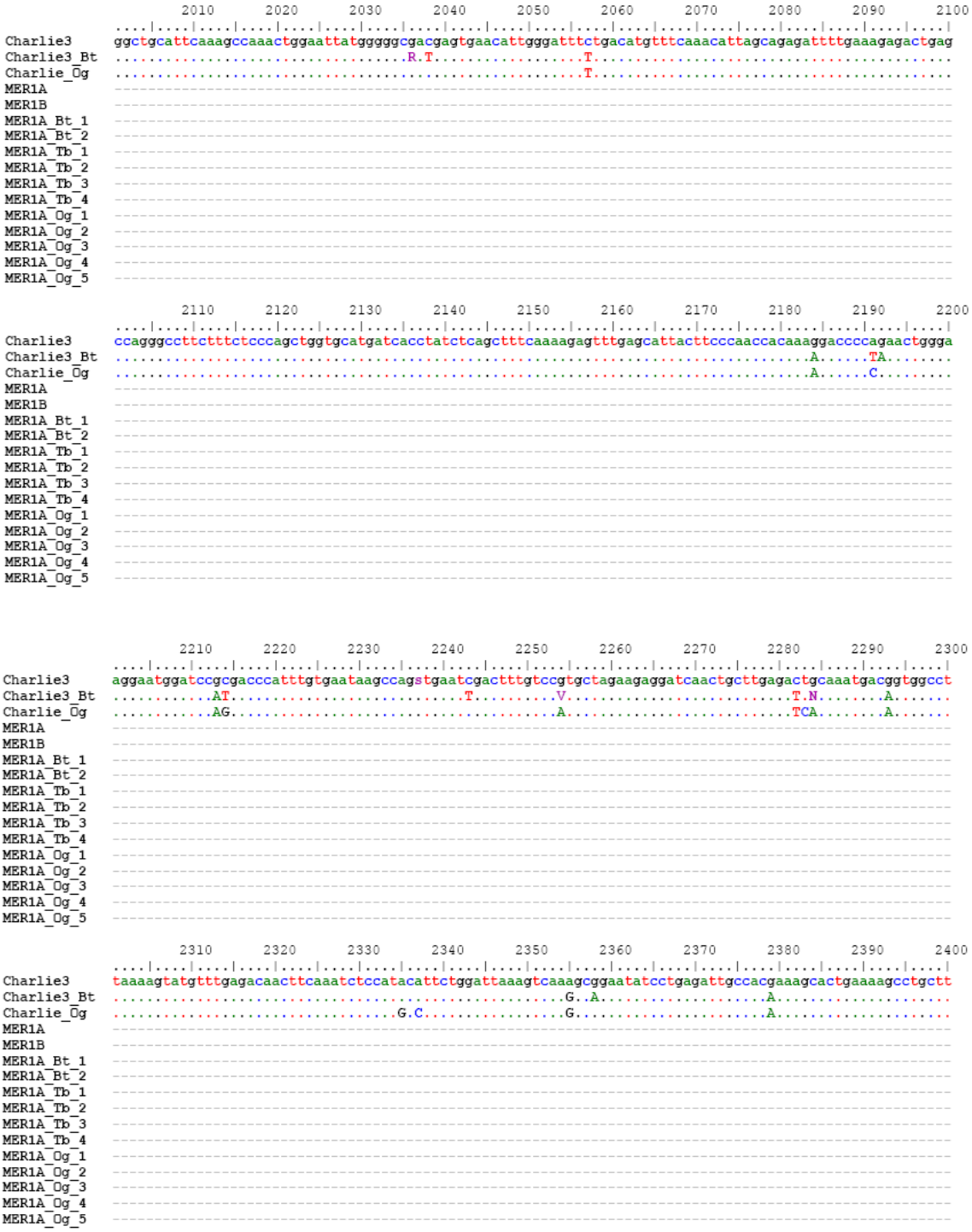


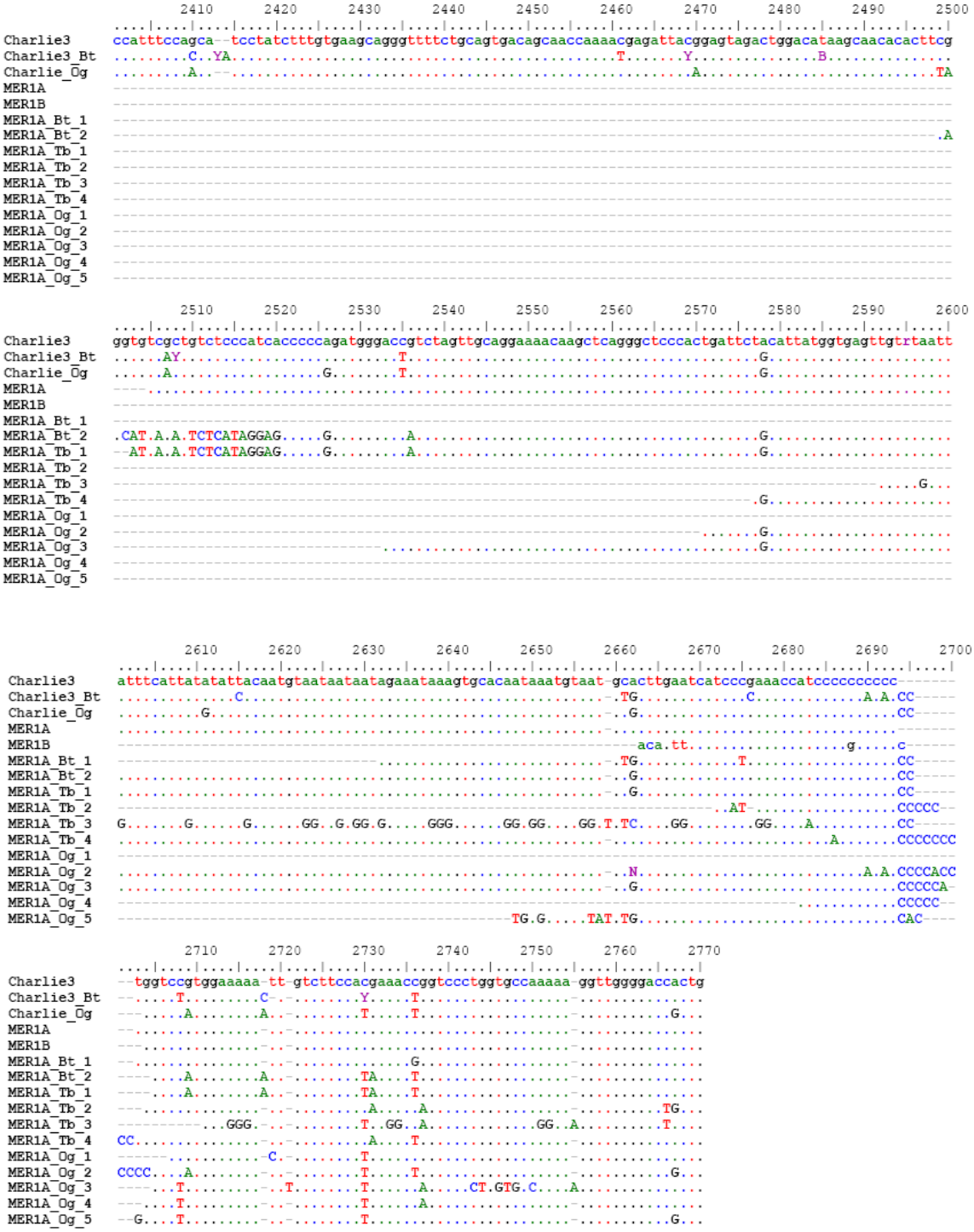

410 420 430 440 450 460 470 480 490 500
Charlie3 agactcatatcaaaaccctatcagtgagtggcaagtgacaattaagctgcatctggtggcaggctttatagtggcaagtgagttgatgtacttcaattgt
Charlie3_Bt
Charlie_Og
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
510 520 530 540 550 560 570 580 590 600
Charlie3 acagctgcatctggtggcaggctttaagtcagaat ccgacacttattttagtccgcgcgtggcctgctcattattttattta
Charlie3_Bt
Charlie_Og
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
610 620 630 640 650 660 670 680 690 700
Charlie3 ccacttccatccgcgcctctttcccgcactgcgcacttgtctcagtcacagttttggtaagcccacaagctaaccctagccaaaatgagtaaaaaacaaa
Charlie3_Bt
Charlie_Og
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
710 720 730 740 750 760 770 780 790 800
Charlie3 cgtcactggagagcttctttgaaaaggggggaaagacccaatgatgagacagcagaagactctaagactgccaamaaaaanaangctgcatttaaaaga
Charlie3_Bt AAT
Charlie_Og
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5


Charlie3
Charlie3_Bt
Charlie_Og
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5

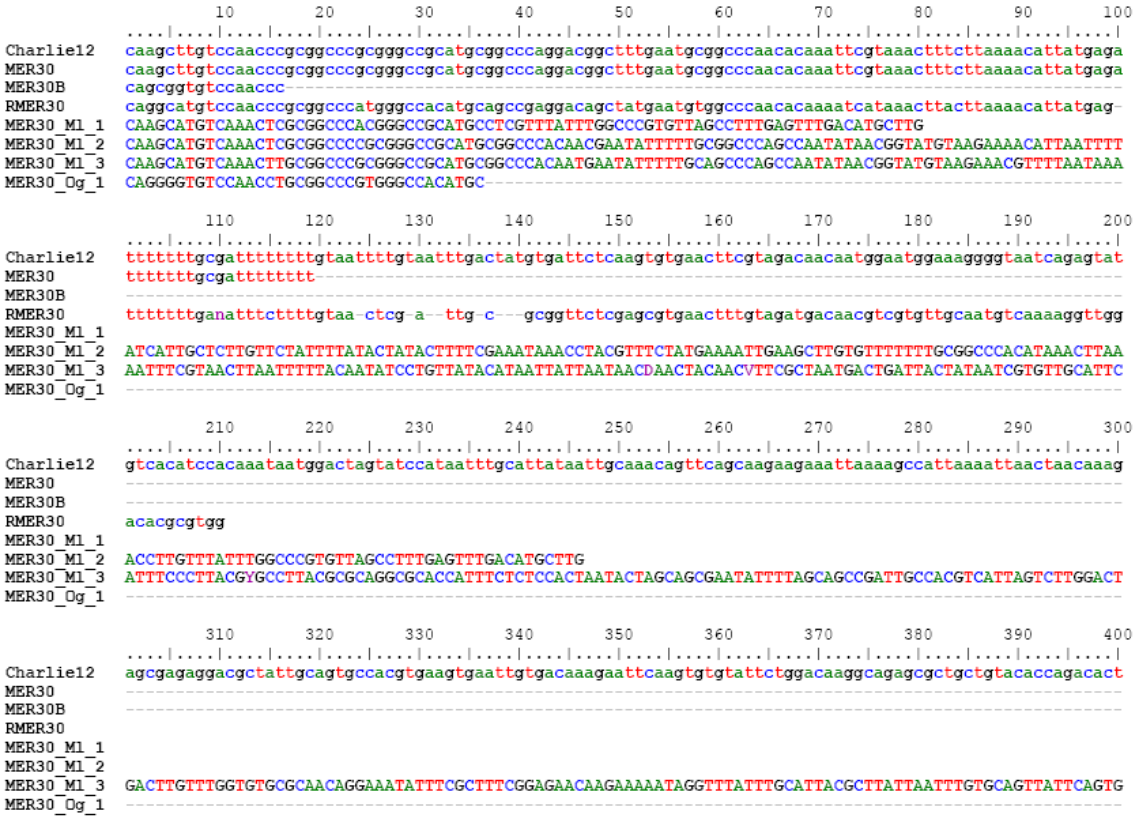

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
Charlie3 tgttcmtctttcakctttcagccaccacaacmagacgaactggtgaaacagcagaggatattgaggcacagttgttagagaggatkaatgagtcaccgtg
Charlie3_Bt ..........TG...AG-----.....T...T...T......A.....T.....................A.....................T...
Charlie_Og ..........TG...AG-----.....T...T.........A.....T.....................A.....................A..
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
Charlie3 gtatgcaatccaggttgamaagtctaccagtgttaacaacaakgmaacaakgcttgcttttgtgkgatatatttttcaggaggatgcgcawgaggatatg
Charlie3_Bt ...H.............BR.........A....G.....................T.......C.....................T.............
Charlie_Og ...C.....................GA....G.....................T.......C.....................T.............
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
Charlie3 ttatgtgcacttttgttgccaaccaacaccacagctgcagaattattcaagtctttgaatgattacatatcaggaaaactgaattggtcattttgtgtcg
Charlie3_Bt .....C.....
Charlie_Og .....C.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
Charlie3 gtatatgcacagacagagtggctgccatgactggatggctttctggtttcactactcgggtcaaagaggtcgcttctaaatgtgagtctacacactgcgt
Charlie3_Bt .........TG..TG...C.....A.....G.....T..
Charlie_Og .........TG...G.....C.....A.....G.....G.....T..
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5

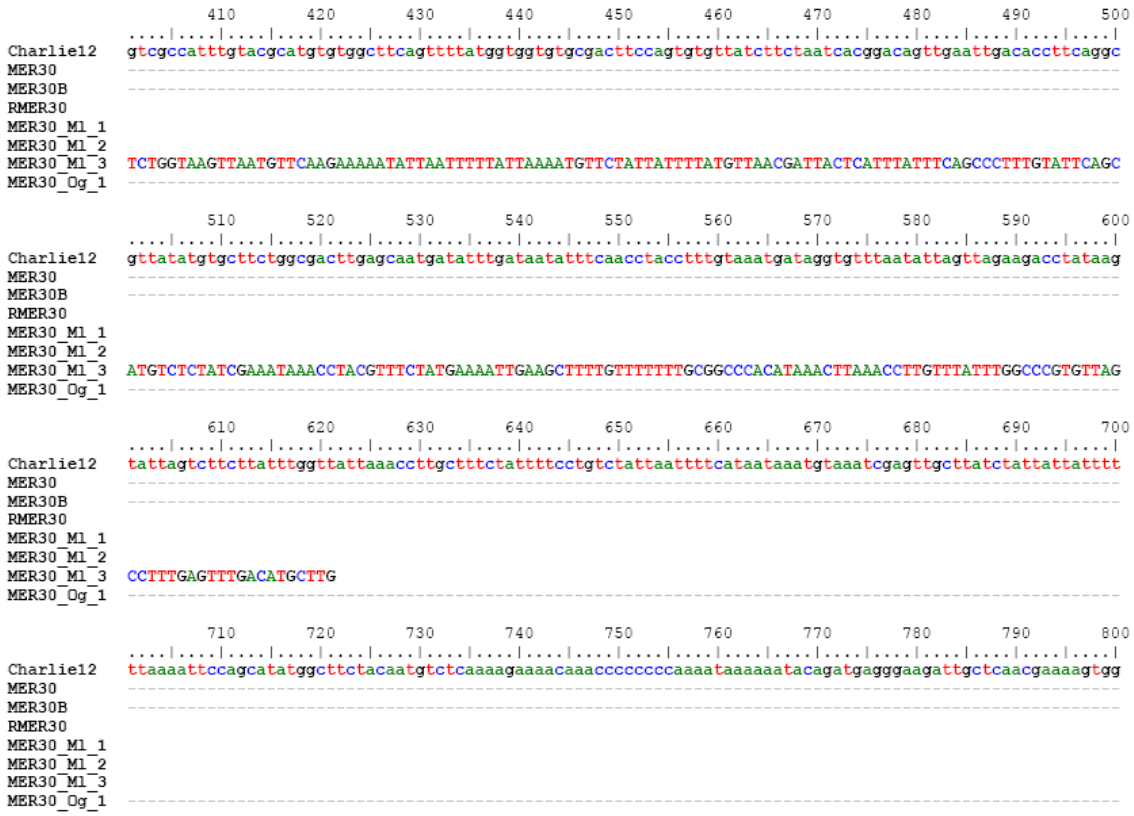

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
Charlie3 catccatagagaaatgctggctagctggaaaatgtcacctgaacttagcmacgttttgcaggatgtsmttaaaattatcaaccacattwaagtacatgcc
Charlie3_Bt ..........A.....A..GT.....
Charlie_Og ..........A.....A.A...T.....G....G.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
Charlie3 cttaactcacatctgttcgcgcagctctgtgaggagatggacgcagmgcacacacgtcttctcttatacacagaagtgagatggctttctaaaggtagat
Charlie3_Bt .....TG.....AT.....T.....A.....R.....
Charlie_Og .....G.....A.....A.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
Charlie3 cactggccagagtttttgagttacgagagccgctccagagatttcttttagaaaaacagtcaccactggcagcacatttcagtgacatagaatgggtcgc
Charlie3_Bt .....T.....C.....
Charlie_Og .....T.....T.....C.....T..
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
Charlie3 aaaacttgcttacttgtgtgacatattcaacctgctcaacgaactcaatctgtcacttcaggggagaatgacaactgtgttcaagtcggcagataaagt
Charlie3_Bt .....T.....GT.....A.....
Charlie_Og .....-.....T.....CA.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5


2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
Charlie3 ggctgcattcaaagccaaactggaattatgggggcgacgagtgaacattgggatttctgacatgtttcaaacattagcagagattttgaaagagactgag
Charlie3_Bt .....R.T.....T.....
Charlie_Og .....T.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
Charlie3 ccagggccttctttctcccagctggtgcatgatcacctatctcagctttcaaaagagtttgagcattacttcccaaccacaaaggaccccagaactgggga
Charlie3_Bt .....A.....TA.....
Charlie_Og .....A.....C.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
Charlie3 aggaatggatccgcgacccatttgtgaataagccagstgaatcgactttgtccgtgctagaagaggatcaactgcttgagactgcaaatgacggtggcct
Charlie3_Bt .....AT.....T.....V.....T.N.....A.....
Charlie_Og .....AG.....A.....TCA.....A.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
Charlie3 taaaagtatgtttgagacaacttcaaatctccatacattctggattaaagtcaaagcggaatatcctgagattgccacgaaagcactgaaaagcctgctt
Charlie3_Bt .....G..A.....A.....
Charlie_Og .....G.C.....G.....A.....
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5


2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
Charlie3 ccatttccagca--tcctatctttgtgaagcagggttttctgcagtgacagcaaccaaaaacgagattacggagtagactggacataagcaacacacttcg
Charlie3_Bt .....C..YA.....T.....Y.....B.....
Charlie_Og .....A..--.....A.....TA
MER1A
MER1B
MER1A_Bt_1
MER1A_Bt_2 .A
MER1A_Tb_1
MER1A_Tb_2
MER1A_Tb_3
MER1A_Tb_4
MER1A_Og_1
MER1A_Og_2
MER1A_Og_3
MER1A_Og_4
MER1A_Og_5
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
Charlie3 ggtgtcgctgtctcccatcaccccagatgggaccgtctagttgcaggaaaacaagctcagggctcccactgattctacattatggtgagttgtrtaatt
Charlie3_Bt ......AY.....T.....G.....
Charlie_Og ......A.....G.....T.....G.....
MER1A ----.....
MER1B
MER1A_Bt_1
MER1A_Bt_2 .CAT.A.A.TCTCATAGGAG.....G.....A.....G.....
MER1A_Tb_1 --AT.A.A.TCTCATAGGAG.....G.....A.....G.....
MER1A_Tb_2
MER1A_Tb_3 .....G...
MER1A_Tb_4 .G.....
MER1A_Og_1
MER1A_Og_2 .....G.....
MER1A_Og_3 .....G.....
MER1A_Og_4
MER1A_Og_5
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
Charlie3 atttcattatatattacaatgtaataataatagaaataaagtgcacaataaatgtaat-gcacttgaatcatcccgaaaccatccccccccc------
Charlie3_Bt .....C.....-.TG.....C.....A.A.CC-----
Charlie_Og .....G.....-..G.....CC-----
MER1A .....-----
MER1B -----aca.tt.....g.....c------
MER1A_Bt_1 -----.....-.TG.....T.....CC-----
MER1A_Bt_2 .....-..G.....CC-----
MER1A_Tb_1 .....-..G.....CC-----
MER1A_Tb_2 -----..AT-.....CCCCC--
MER1A_Tb_3 G.....G.....G.....GG..G.GG.G.....GGG.....GG.GG....GG.T.TC....GG.....GG....A.....CC-----
MER1A_Tb_4 .....-.....A.....CCCCCCC
MER1A_Og_1 -----
MER1A_Og_2 .....-..N.....A.A.CCCCACC
MER1A_Og_3 .....-..G.....CCCCCA-
MER1A_Og_4 -----.....CCCCC--
MER1A_Og_5 -----TG.G.....TAT.TG.....CAC----
2710 2720 2730 2740 2750 2760 2770
Charlie3 --tggtccgtggaaaaa-tt-gtcttccacgaaaccggtccctggtgccaaaaa-ggttggggaccactg
Charlie3_Bt --.....T.....C.-.....Y.....T.....-.....
Charlie_Og --.....A.....A.-.....T.....T.....-.....G...
MER1A --.....-..-.....-.....
MER1B --.....-..-.....-.....
MER1A_Bt_1 --.....-..-.....G.....-.....
MER1A_Bt_2 ----....A.....A.-.....TA....T.....-.....
MER1A_Tb_1 ----....A.....A..-.....TA....T.....-.....
MER1A_Tb_2 --.....-..-.....A.....A.....-.....TG...
MER1A_Tb_3 ----------...GGG.-..-.....T..GG..A.....GG..A.....T....
MER1A_Tb_4 CC.....-..-.....A....T.....-.....
MER1A_Og_1 -----.....C.-.....T.....-.....
MER1A_Og_2 CCCC....A.....-..-.....T.....T.....-.....G...
MER1A_Og_3 ----...T.....-..T.....T.....A.....CT.GTG.C....A.....
MER1A_Og_4 --....T.....-..-.....T.....A.....-.....
MER1A_Og_5 --G....T.....-..-.....T.....-.....G...

Charlie12 Tribe



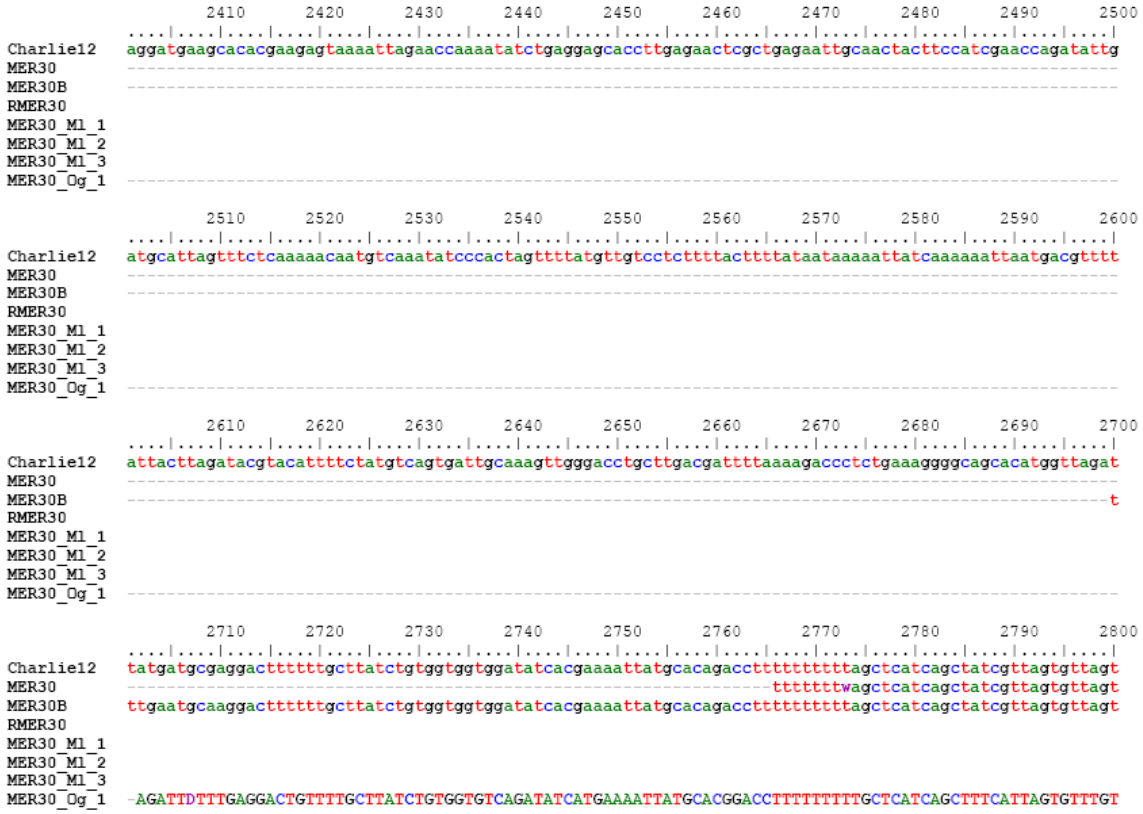

410 420 430 440 450 460 470 480 490 500
Charlie12 gtcgccatttgtacgcatgtgtggcttcagttttatggtggtgtgcgacttccagtgtgttatcttctaatcacggacagttgaattgacaccttcaggc
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3 TCTGGTAAGTTAATGTTCAAGAAAAATATTAATTTTTATTAAAATGTTCTATTATTTTATGTTAACGATTACTCATTTATTTCAGCCCTTTGTATTCAGC
MER30_Og_1
510 520 530 540 550 560 570 580 590 600
Charlie12 gttatatgtgcttctggcgacttgagcaatgatatttgataatatttcaacctacctttgtaaatgataggtgtttaatattagttagaagacctataag
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3 ATGTCTCTATCGAAATAAACCTACGTTTCTATGAAAATTGAAGCTTTTGTTTTTTTGCGGCCCACATAAACTTAAACCTTGTTTATTTGGCCCGTGTTAG
MER30_Og_1
610 620 630 640 650 660 670 680 690 700
Charlie12 tattagtcttcttatttggttattaaaccttgctttctattttcctgtctattaatttttcataataaatgtaaatcgagttgcttatctattattatttt
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3 CCTTTGAGTTTGACATGCTTG
MER30_Og_1
710 720 730 740 750 760 770 780 790 800
Charlie12 ttaaaattccagcatatggcttctacaatgtctcaaaagaaaacaaacccccccccaaaataaaaaatacagatgagggaagattgctcaacgaaaagtgg
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1

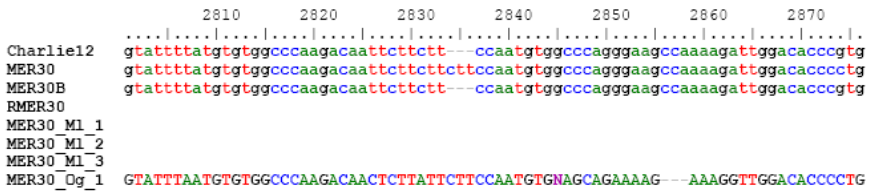

810 820 830 840 850 860 870 880 890 900
Charlie12 acagatgactactttttgtcaaggcaaatagtaaggcactctgcttgatttgtagggaatttgtgccagtttcaaagactataatttgaaaaggcatta
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
910 920 930 940 950 960 970 980 990 1000
Charlie12 tatgcaaagacgtgctgccaaatttggtgcgtatcaaggaatgtgtcgtaaggacaaaaatagcagaactgaaaaaatgtctgtcttcacaaaaaaaatt
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
Charlie12 ttttttaaagttgcaactcaaacagtctattgtaaaagctagttatatgatagcaaatttaatagcaaaaagcaaaactatttacagatggtgagtttat
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
Charlie12 taagcaacgtatgggaggcatggcatatatcatttgccctgataaaaaagaagatatctctaaaatcagtttgtcttgccggaatatagccaggtgaatt
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
Charlie12 ggagaaattggaaagtctatgaaaagagcgtaaaactgctaattttaaattttgtgctttggcgatggatgaaagcactgatgctacacatatggcacaa
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
Charlie12 cttgccatttttattagaggcattgatgacgaatagaatgtcatcattatataaagccatataatgaagaaataataaacccatataatgaagaaatggc
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
Charlie12 ttttttagtgccattaaaaaacagagtaaatcaagagatttatacgaagtagtaaaagatgcattaaagcaattttctttgtgcgttgtgaacatacctg
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
Charlie12 gtatagttactgatgatgcccctgcgatggtacgtaaaagagagggagttgtaaaattaatagaaaatgatgcagttgccgcctgaaactcacttttgat
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1

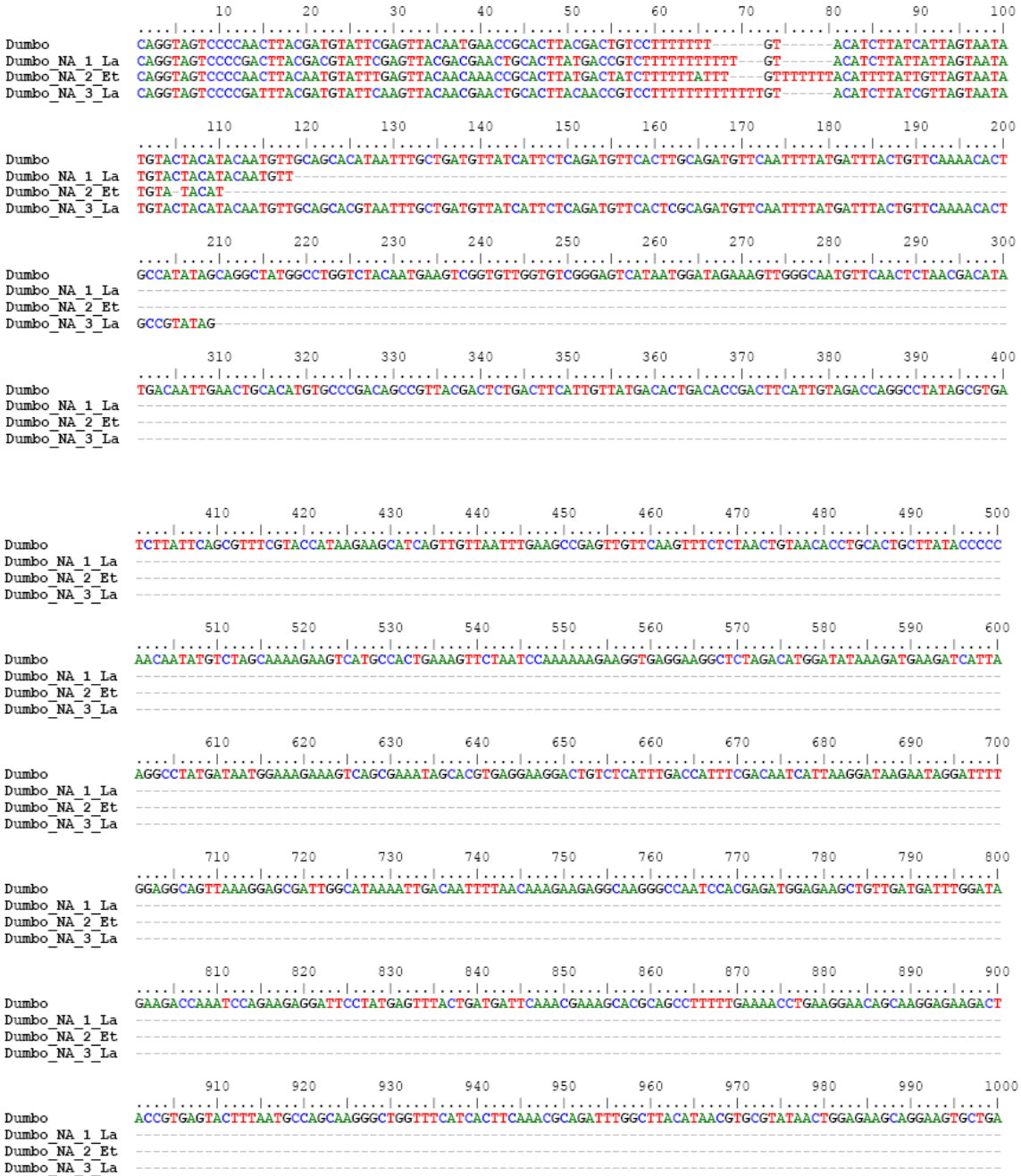

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
Charlie12 gatgtgtcattgtatagtacatcaagaaaatttatgcacaaaagctttaaaaatggataacatcatgcaaattgtcatcaaggctgtgaatttcatagggg
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
Charlie12 gccaagagattgaatcattgccaattccaggaattccttaaaagtatggatgctgactatagcaacatcatttacttttcggaagtaaagtcgagacaga
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
Charlie12 tgttgaaaagattttatgatttgcgacatgaaatcgagttatttatggtatcaaaaacaaaatttgtgccagaacttgatgacgaaaactggcttacaga
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
Charlie12 tttagcatttttagtggatttgaccactcatttaaatgagttaaacatgaatcttcaaggtgaaaaccaacttctcaatacaatgtttcaaaccataaca
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1


2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
Charlie12 gtgttccaaatacaattgaaatttatggcaagctaaaattaaggcaaacagttttacggatttcaacacatttgctaaacacgggcttgtcaacagcaaaa
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
Charlie12 agtattctgccttgctttttgatttgataaaggaatttgaaaacaggttttaagatttctggaaaaataatcaatattttggtatagttgcaactccatt
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
Charlie12 ttcagccaacataaatatgttacctgcgaatgcatacagctgcaatgtgacattcaacttaaagaaaaatctcatcaggcttctttcctggactttgtaa
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
Charlie12 gacctatcttcccagagacaaatatccctcgcttcacagtcatgccttactcatgtcgtcggttttggcagcacctgcgtttgtgagcaactgttttca
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1


2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
Charlie12 aggatgaagcacacgaagagtaaaattagaaccaaaatatctgaggagcaccttgagaactcgctgagaattgcaactacttccatcgaaccagatattg
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
Charlie12 atgcattagtttctcaaaaacaatgtcaaatatcccactagttttatgttgtcctcttttacttttataataaaaattatcaaaaaattaatgacgtttt
MER30
MER30B
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
Charlie12 attacttagatacgtacattttctatgtcagtgattgcaaagttgggacctgcttgacgattttaaaagaccctctgaaaggggcagcacatggttagat
MER30
MER30B t
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
Charlie12 tatgatgcgaggacttttttgcttatctgtggtggtggatatcacgaaaattatgcacagacctttttttttttagctcatcagctatcgttagtgttagt
MER30 ttttttttwagctcatcagctatcgttagtgttagt
MER30B ttgaatgcaaggacttttttgcttatctgtggtggtggatatcacgaaaattatgcacagacctttttttttttagctcatcagctatcgttagtgttagt
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1 -AGATTDTTTGAGGACTGTTTTGCTTATCTGTGGTGTCAGATATCATGAAAATTATGCACGGACCTTTTTTTTTTGCTCATCAGCTTTCATTAGTGTTTGT


2810 2820 2830 2840 2850 2860 2870
Charlie12 gtattttatgtgtggcccaagacaattcttctt---ccaatgtggcccagggaagccaaaagattggacacccgtg
MER30 gtattttatgtgtggcccaagacaattcttcttcttccaatgtggcccagggaagccaaaagattggacacccctg
MER30B gtattttatgtgtggcccaagacaattcttctt---ccaatgtggcccagggaagccaaaagattggacacccgtg
RMER30
MER30_Ml_1
MER30_Ml_2
MER30_Ml_3
MER30_Og_1 GTATTTAATGTGTGGCCCAAGACAACTCTTATTCTTCCAATGTGNAGCAGAAAAG---AAAGGTTGGACACCCCTG

Dumbo Tribe

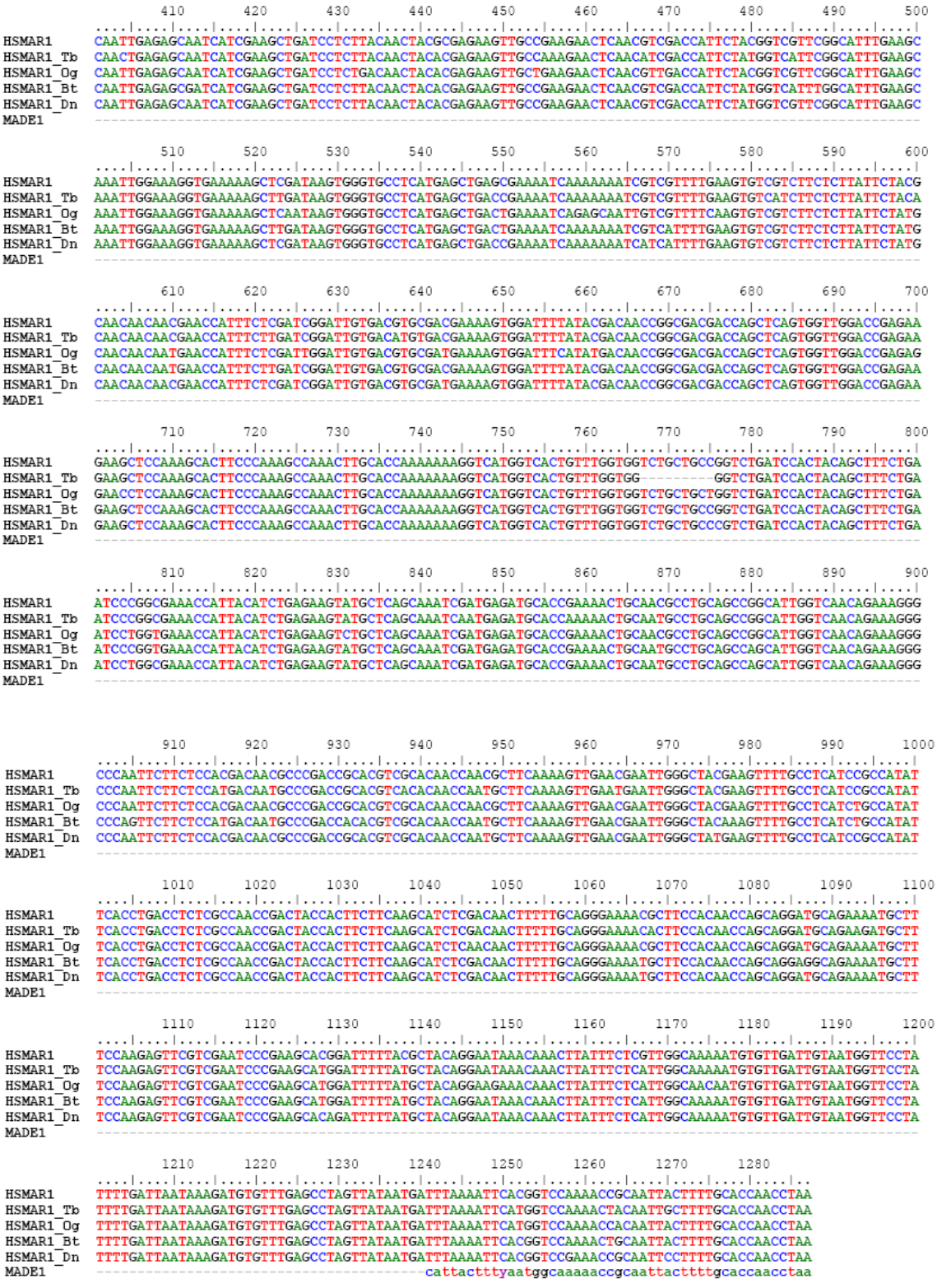

10 20 30 40 50 60 70 80 90 100
Dumbo CAGGTAGTCCCCAACTTACGATGTATTCGAGTTACAATGAACCGCACTTACGACTGTCCTTTTTTT------GT------ACATCTTATCATTAGTAATA
Dumbo_NA_1_La CAGGTAGTCCCCGACTTACGACGTATTCGAGTTACGACGAACTGCACTTATGACCGTCTTTTTTTTTTT---GT------ACATCTTATTATTAGTAATA
Dumbo_NA_2_Et CAGGTAGTCCCCAACTTACAATGTATTTGAGTTACAACAAACCGCACTTATGACTATCTTTTTTATTT----GTTTTTTTACATTTTATTGTTAGTAATA
Dumbo_NA_3_La CAGGTAGTCCCCGATTTACGATGTATTCAAGTTACAACGAACTGCACTTACAACCGTCCTTTTTTTTTTTTTGT------ACATCTTATCGTTAGTAATA
110 120 130 140 150 160 170 180 190 200
Dumbo TGTACTACATACAATGTTGCAGCACATAATTTGCTGATGTTATCATTCTCAGATGTTCACTTGCAGATGTTCAATTTTATGATTTACTGTTCAAAACACT
Dumbo_NA_1_La TGTACTACATACAATGTT
Dumbo_NA_2_Et TGTA-TACAT
Dumbo_NA_3_La TGTACTACATACAATGTTGCAGCACGTAATTTGCTGATGTTATCATTCTCAGATGTTCACTCGCAGATGTTCAATTTTATGATTTACTGTTCAAAACACT
210 220 230 240 250 260 270 280 290 300
Dumbo GCCATATAGCAGGCTATGGCCTGGTCTACAATGAAGTCGGTGTTGGTGTCGGGAGTCATAATGGATAGAAAGTTGGGCAATGTTCAACTCTAACGACATA
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La GCCGTATAG
310 320 330 340 350 360 370 380 390 400
Dumbo TGACAATTGAACTGCACATGTGCCCGACAGCCGTTACGACTCTGACTTCATTGTTATGACACTGACACCGACTTCATTGTAGACCAGGCCTATAGCGTGA
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La
410 420 430 440 450 460 470 480 490 500
Dumbo TCTTATTCAGCGTTTCGTACCATAAGAAGCATCAGTTGTTAATTTGAAGCCGAGTTGTTCAAGTTTCTCTAACTGTAACACCTGCACTGCTTATACCCCC
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La
510 520 530 540 550 560 570 580 590 600
Dumbo AACAATATGTCTAGCAAAAGAAGTCATGCCACTGAAAGTTCTAATCCAAAAAAGAAGGTGAGGAAGGCTCTAGACATGGATATAAAGATGAAGATCATTA
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La
610 620 630 640 650 660 670 680 690 700
Dumbo AGGCCTATGATAATGGAAAGAAAGTCAGCGAAATAGCACGTGAGGAAGGACTGTCTCATTTGACCATTTCGACAATCATTAAGGATAAGAATAGGATTTT
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La
710 720 730 740 750 760 770 780 790 800
Dumbo GGAGGCAGTTAAAGGAGCGATTGGCATAAAATTGACAATTTTAACAAAGAAGAGGCAAGGGCCAATCCACGAGATGGAGAAGCTGTTGATGATTTGGATA
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La
810 820 830 840 850 860 870 880 890 900
Dumbo GAAGACCAAATCCAGAAGAGGATTCCTATGAGTTTACTGATGATTCAAACGAAAGCACGCAGCCTTTTTGAAAACCTGAAGGAACAGCAAGGAGAAGACT
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La
910 920 930 940 950 960 970 980 990 1000
Dumbo ACCGTGAGTACTTTAATGCCAGCAAGGGCTGGTTTCATCACTTCAAACGCAGATTTGGCTTACATAACGTGCGTATAACTGGAGAAGCAGGAAGTGCTGA
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La


1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
Dumbo TGTTGAAGCTGCCAAAAAATTCAAGAATGAGCTGGACAAGATGGAAGAGGAAGGAGAGAAGTTAAAAAATTTTACGGTGAAAGGATTAGCTGGAGTTTTT
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La ATTTAGTTT
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
Dumbo TTTAAACTAAATCAGGGGCTTCAGTTGCTTGAAAACATGGACCCAAACACTGATTGCTTTGCTAAAGTCGATAGGCAGATTTGGGAAGCAGTGAGATGTT
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La TCTAAACTAAATCAGGGGCTTCAGTTGCTTGAAAACATGGACCCAAATGCTGATCACTTTGCTAAAGTTGATAGGCAGATTTGGGAAGCAGTGAGATGTT
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
Dumbo ACTATGAAATATATGAAGAAAAAAAGAAAAGGACCATACAGACAACTCTCACTAATTTTCTGACTAGAAACACCATCCCCATTCCCAGGGATAATCCAGA
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La ACCATGAAATATATGAAGAAAAAAAGAAAAGGACCATACAGACAACTCTCACTAATTTTCTGACTAGAAATGCCATCCCCATTCCCAGGGATAATCCAGA
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
Dumbo TGATCCAGAACCTTCCACAAGTGGAGTTATTACTGCAACTGACAAAATGCCAATGTCGTCCAACTCTTCTTTATCATCAGAGTAAATTTTT ATGATTTG
Dumbo_NA_1_La
Dumbo_NA_2_Et
Dumbo_NA_3_La TGATCCAGAACCTTCCACAAGTGGAGTTATTACCGCAACTGACAAAATGCCAACGTCGTCCAACTCTTCTTCATCGTTGGAGTAAATTTTTTATGATTTG
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
Dumbo TGTATTTTTTTAGTGTATGTATGTATGTATTAAAATATATCTGTAAGTTTATATGTAATGTTTCCAACCCCCAAAGACAAA TAAAGATCAGATTTATAA
Dumbo_NA_1_La AAAATATATCTGTAAGTTTATATGTAATGTTTCCAACCCCCAAAGACAAA TAAAGATCAGATTTATAA
Dumbo_NA_2_Et TACATATATTTAATTATATCTTTAAGTTTATATGTAATGTTTCCAACCCCCAAAGACAAA TAAAGATCAGATTTATAA
Dumbo_NA_3_La TGTATTTTTTTA TGTATGTATGTATGTATTAAAATATATCTGTAAGTTTATATGTAATGTTTCCAACCCCCAAAGACAAAATAAAGATCGGATTTATAA
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
Dumbo A GATACTGATAATAAAAGGCAATAATAATGAAAACTAAAAAAAAAAAAATGAGATATTTGACTTACATCAGAACCGACTTACGACAGAGTCGTCAGAATG
Dumbo_NA_1_La A GATACTGATAATAAAAGGCAATAATAATGAAAACTAAAA TGAGGTATTTGACTTACATCAGAACCAACTTATGATGGAGTCGTCAGAATG
Dumbo_NA_2_Et AAGATACTGATAATAAAAGGCAATAATAATGAAAACT TAAAAGAGGTATTCGACTTACGTCAGAACCGATTTATGACGGAGTCGTCGGAATA
Dumbo_NA_3_La A GATACTGATAATAAAAGGCAATAATAATGAAAACTAAAAAAAAAA GAGGTATTCGACTTACGTCAGAACCGACTTACAATGGAGTTGTCGGAATG

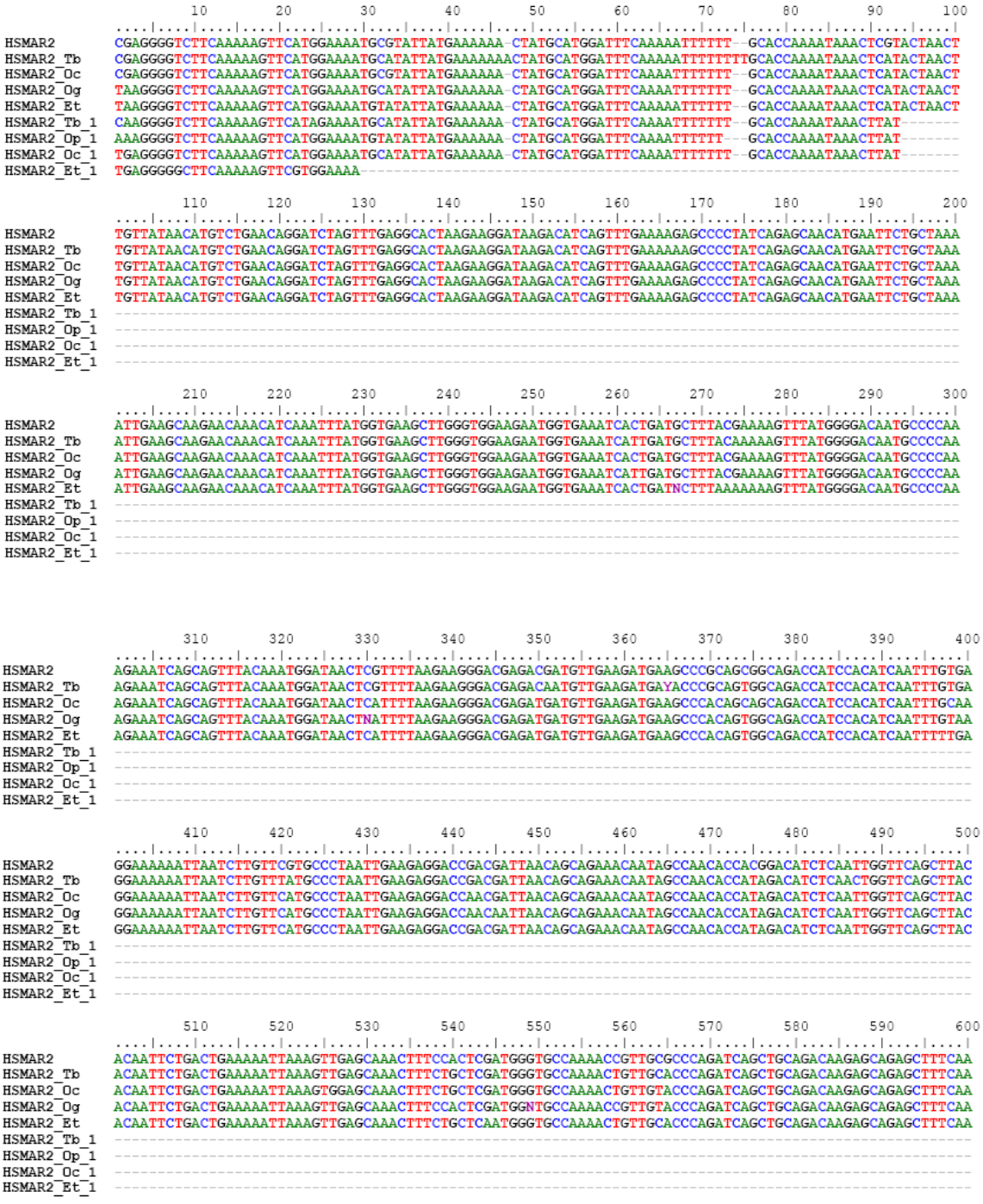
```
                    1610      1620
               ....|....|....|....|....
Dumbo          GAACCCTGTCGTAAGTTGGGGACTACCTG
Dumbo_NA_1_La  GAACCCTGTCATAAGTCAGGGACTACCTG
Dumbo_NA_2_Et  GAACCCCGTCATAAGTTGGGGACTACCTG
Dumbo_NA_3_La  GAACCCCATCGTAAGTCGGGGACTACCTG
```

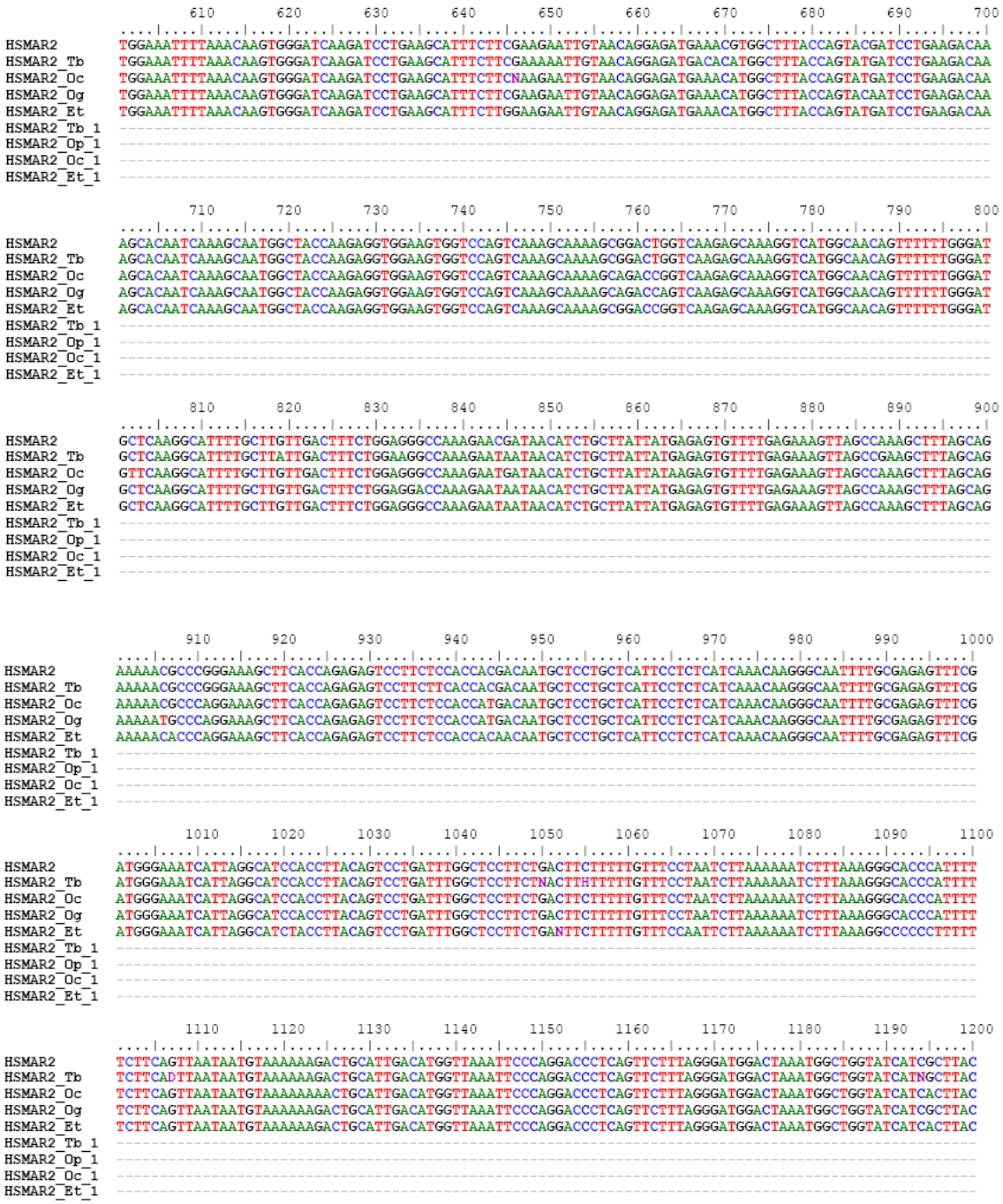
HSMAR1 Tribe

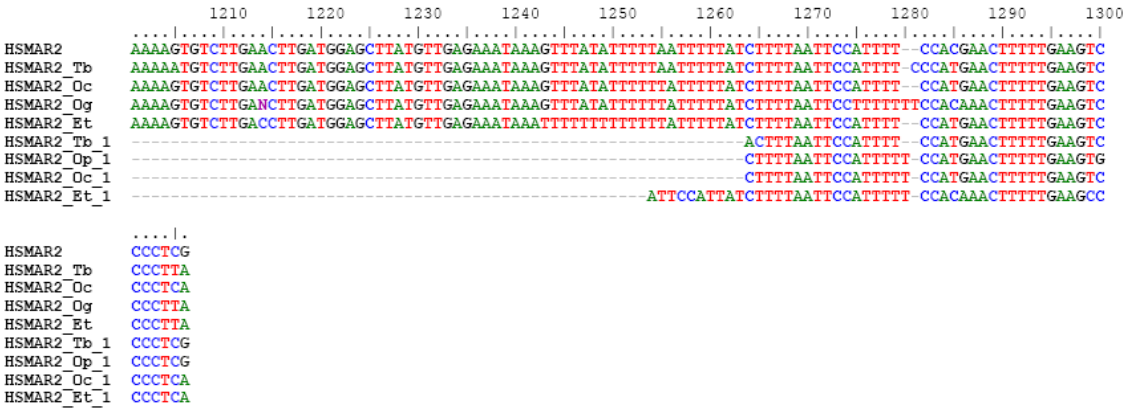
```
                10        20        30        40        50        60        70        80        90        100
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HSMAR1         TTAGGTTGGTGCAAAAGTAATTGCGGTTTTTGCATTGTTGGAAATTTGCCGTTTGATATTGGAATACATTCTTAAATAAATGTGGTTATGTTATACATCAT
HSMAR1_Tb      TTAGGTTGGTGCAAAAGTAATTGCGGTTTT-GCATTGTTGAAATTTGCCATTTGATATTGGA-TACATTCTTAAATAAATGTGGTTATGTTATACATCAT
HSMAR1_Og      TTAGGTTGGTGCAAAAGTAATTGCGGTTTTTGCATTGTTGAAATTTGCCATTTGATACTGGAATACATTCTTAAATAAATGTGGTTATGTTATACATCAT
HSMAR1_Bt      TTAGGTTGGTGCAAAAGTAATTGCGGTTTT-GCATTGTTGAACTTTGCCATTTGATATTGGAATACATTCTTAAATAAATGTGGTTATGTTATACATCAT
HSMAR1_Dn      TTAGGTTGGTGCAAAAGTAATTGCGGTTTT-GCATTGTTGAAATTTGCCATTTGATATTGGCATACATTCTTAAATAAATGTGGTTATGTTATACATCAT
MADE1          ttaggttggtgcaaaagtaattgcggttttgc----------------------------------------------------------------

                110       120       130       140       150       160       170       180       190       200
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HSMAR1         TTTAATGCGCATTTCTCGCTTTACGTTTTTTGCTAATGACTTATTACTTGCTGTTTATTTTATGTTTATTTTAGACTATGGAAATGATGTTAGACAAAA
HSMAR1_Tb      TTTAATGCACATTTCTTGCTTT---TTTTTTGCTAATGACTTATTACTTGCTGTTTATTTTATATGTATTTTAGACTATGGAAATGATGTTAGACAAAA
HSMAR1_Og      TTTAATGCACATTTCTCGCTTTATGTTTTTTTGCTAATGACTTATTACTTGCTGTTTATTTTATATTTATTTTGGACTATGGAAATGATGTTAGACAAAA
HSMAR1_Bt      TTTAATGTACATTTCTCGCTTTATGTTTTTTTGCTAATGACTTATTACTTGCTGTTTATTTTATATTTATTTTAGACTATGGAAATGATGTTAGACAAAA
HSMAR1_Dn      TTTAATGTGCATTTCTCGCTTTATGTTTTTTTGCTAATGACTTATTACTTGCTGTTTATTTTATATTTATTTTAGACTATGGAAATGATGTTAGACAAAA
MADE1          ----------------------------------------------------------------------------------------------------

                210       220       230       240       250       260       270       280       290       300
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HSMAR1         AGCAAATTCGAGCGATTTTCTTATTCGAGTTCAAAATGGGTCGTAAAGCGGCGGAGACAACTCGCAACATCAACAACGCATTTGGCCCAGGAACTGCTAA
HSMAR1_Tb      AGCAAATTTGAGTGATTTTCTTATTCGAGTTCAAAATGGGTCGTAAAGCAGCGGAGACAACTCGCAACATCAACAACGCAATTGGCCCAGGAACTGCTAA
HSMAR1_Og      AGCAAATTCGAGCGATTTTCTTATTCGAGTTCAAAATGGGTCGTAAAGCAGCAGAGACAACTCGCAACATCAACAACGCATTTGGCCCAGGAACTGCTAA
HSMAR1_Bt      AGCAAATTCGAGCGATTTTCTTATTCGAGTTCAAAATGGGTCATAAAGCAGCAGAGACAACTCACAACATCAACAATGCATTTGGCCCAGGAACTGCTAA
HSMAR1_Dn      AGCAAATTCGAGCGATTTTCTTATTTGAGTTCAAAATGGGTCGTAAAGCAGCAGAGACAACTTGCAACATCAACAACGCATTTGGCCCAGGAACTGCTAA
MADE1          ----------------------------------------------------------------------------------------------------

                310       320       330       340       350       360       370       380       390       400
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
HSMAR1         CGAACGTACAGTGCAGTGGTGGTTCAAGAAGTTTTGCAAAGGAGACGAGAGCCTTGAAGATGAGGAGCGTAGTGGCCGGCCATCGGAAGTTGACAACGAC
HSMAR1_Tb      CGAACGTACAGTGCAGTGGTGGTTCAAGAAGTTTTGCAAAGGAGACGAGAGCCTTGAAGATGAGGAGCATAGTGGCCGGCCATCGGAAGTTGACAACGAC
HSMAR1_Og      TGAACGTACAGTGCAGTGGTGGTTCAAGAAGTTTTGCAAAGGAGACGAGAGCCTTGAAGATGAGGAGCATAGTGGCCGGCCATCGGAAGTTGACAACGAC
HSMAR1_Bt      TGAACATACAGTGCAGTGGTGGTTCAAGAAGTTTTGCAAAGGAGACGAGAGCCTTGAAGATGAGGAGCGCAGTGGCTGGCCATCGGAAGTTGACAACGAC
HSMAR1_Dn      CGAACGTACAGTGCAGTGGTGGTTCAAGAAGTTTTGCAAAGGAGACGAGAGCCTTGAAGATGAGGAGCGTAGTGGCCGGCCATCGGAAGTTGACAACGAC
MADE1          ----------------------------------------------------------------------------------------------------
```


410 420 430 440 450 460 470 480 490 500
HSMAR1 CAATTGAGAGCAATCATCGAAGCTGATCCTCTTACAACTACGCGAGAAGTTGCCGAAGAACTCAACGTCGACCATTCTACGGTCGTTCGGCATTTGAAGC
HSMAR1_Tb CAACTGAGAGCAATCATCGAAGCTGATCCTCTTACAACTACACGAGAAGTTGCCAAAGAACTCAACATCGACCATTCTATGGTCATTCGGCATTTGAAGC
HSMAR1_Og CAATTGAGAGCAATCATCGAAGCTGATCCTCTGACAACTACACGAGAAGTTGCTGAAGAACTCAACGTTGACCATTCTACGGTCGTTCGGCATTTGAAGC
HSMAR1_Bt CAATTGAGAGCGATCATCGAAGCTGATCCTCTTACAACTACACGAGAAGTTGCCGAAGAACTCAACGTCGACCATTCTATGGTCATTTGGCATTTGAAGC
HSMAR1_Dn CAATTGAGAGCAATCATCGAAGCTGATCCTCTTACAACTACACGAGAAGTTGCCGAAGAACTCAACGTCGACCATTCTATGGTCGTTCGGCATTTGAAGC
MADE1
510 520 530 540 550 560 570 580 590 600
HSMAR1 AAATTGGAAAGGTGAAAAAGCTCGATAAGTGGGTGCCTCATGAGCTGAGCGAAAATCAAAAAAATCGTCGTTTTGAAGTGTCGTCTTCTCTTATTCTACG
HSMAR1_Tb AAATTGGAAAGGTGAAAAAGCTTGATAAGTGGGTGCCTCATGAGCTGACCGAAAATCAAAAAAATCGTCGTTTTGAAGTGTCATCTTCTCTTATTCTACA
HSMAR1_Og AAATTGGAAAGGTGAAAAAGCTCAATAAGTGGGTGCCTCATGAGCTGACTGAAAATCAGAGCAATTGTCGTTTTCAAGTGTCGTCTTCTCTTATTCTATG
HSMAR1_Bt AAATTGGAAAGGTGAAAAAGCTTGATAAGTGGGTGCCTCATGAGCTGACTGAAAATCAAAAAAATCGTCATTTTGAAGTGTCGTCTTCTCTTATTCTATG
HSMAR1_Dn AAATTGGAAAGGTGAAAAAGCTCGATAAGTGGGTGCCTCATGAGCTGACCGAAAATCAAAAAAATCATCATTTTGAAGTGTCGTCTTCTCTTATTCTATG
MADE1
610 620 630 640 650 660 670 680 690 700
HSMAR1 CAACAACAACGAACCATTTCTCGATCGGATTGTGACGTGCGACGAAAAGTGGATTTTATACGACAACCGGCGACGACCAGCTCAGTGGTTGGACCGAGAA
HSMAR1_Tb CAACAACAACGAACCATTTCTTGATCGGATTGTGACATGTGACGAAAAGTGGATTTTATACGACAACCGGCGACGACCAGCTCAGTGGTTGGACCGAGAA
HSMAR1_Og CAACAACAATGAACCATTTCTCGATTGGATTGTGACGTGCGATGAAAAGTGGATTTCATATGACAACCGGCGACGACCAGCTCAGTGGTTGGACCGAGAG
HSMAR1_Bt CAACAACAATGAACCATTTCTTGATCGGATTGTGACGTGCGACGAAAAGTGGATTTTATACGACAACCGGCGACGACCAGCTCAGTGGTTGGACCGAGAA
HSMAR1_Dn CAACAACAACGAACCATTTCTCGATCGGATTGTGACGTGCGATGAAAAGTGGATTTTATACGACAACCGGCGACGACCAGCTCAGTGGTTGGACCGAGAA
MADE1
710 720 730 740 750 760 770 780 790 800
HSMAR1 GAAGCTCCAAAGCACTTCCCAAAGCCAAACTTGCACCAAAAAAAAGGTCATGGTCACTGTTTGGTGGTCTGCTGCCGGTCTGATCCACTACAGCTTTCTGA
HSMAR1_Tb GAAGCTCCAAAGCACTTCCCAAAGCCAAACTTGCACCAAAAAAAAGGTCATGGTCACTGTTTGGTGG---------GGTCTGATCCACTACAGCTTTCTGA
HSMAR1_Og GAACCTCCAAAGCACTTCCCAAAGCCAAACTTGCACCAAAAAAAAGGTCATGGTCACTGTTTGGTGGTCTGCTGCTGGTCTGATCCACTACAGCTTTCTGA
HSMAR1_Bt GAAGCTCCAAAGCACTTCCCAAAGCCAAACTTGCACCAAAAAAAAGGTCATGGTCACTGTTTGGTGGTCTGCTGCCGGTCTGATCCACTACAGCTTTCTGA
HSMAR1_Dn GAAGCTCCAAAGCACTTCCCAAAGCCAAACTTGCACCAAAAAAAAGGTCATGGTCACTGTTTGGTGGTCTGCTGCCCGTCTGATCCACTACAGCTTTCTGA
MADE1
810 820 830 840 850 860 870 880 890 900
HSMAR1 ATCCCGGCGAAACCATTACATCTGAGAAGTATGCTCAGCAAATCGATGAGATGCACCGAAAACTGCAACGCCTGCAGCCGGCATTGGTCAACAGAAAGGG
HSMAR1_Tb ATCCCGGCGAAACCATTACATCTGAGAAGTATGCTCAGCAAATCAATGAGATGCACCAAAAACTGCAATGCCTGCAGCCGGCATTGGTCAACAGAAAGGG
HSMAR1_Og ATCCTGGTGAAACCATTACATCTGAGAAGTCTGCTCAGCAAATCGATGAGATGCACCGAAAACTGCAACGCCTGCAGCCGGCATTGGTCAACAGAAAGGG
HSMAR1_Bt ATCCCGGTGAAACCATTACATCTGAGAAGTATGCTCAGCAAATCGATGAGATGCACCGAAAACTGCAATGCCTGCAGCCAGCATTGGTCAACAGAAAGGG
HSMAR1_Dn ATCCTGGCGAAACCATTACATCTGAGAAGTATGCTCAGCAAATCGATGAGATGCACCGAAAACTGCAATGCCTGCAGCCAGCATTGGTCAACAGAAAGGG
MADE1
910 920 930 940 950 960 970 980 990 1000
HSMAR1 CCCAATTCTTCTCCACGACAACGCCCGACCGCACGTCGCACAACCAACGCTTCAAAAGTTGAACGAATTGGGCTACGAAGTTTTGCCTCATCCGCCATAT
HSMAR1_Tb CCCAATTCTTCTCCATGACAATGCCCGACCGCACGTCACACAACCAATGCTTCAAAAGTTGAATGAATTGGGCTACGAAGTTTTGCCTCATCCGCCATAT
HSMAR1_Og CCCAATTCTTCTCCACGACAACGCCCGACCGCACGTCGCACAACCAACGCTTCAAAAGTTGAACGAATTGGGCTACGAAGTTTTGCCTCATCTGCCATAT
HSMAR1_Bt CCCAGTTCTTCTCCATGACAATGCCCGACCACACGTCGCACAACCAATGCTTCAAAAGTTGAACGAATTGGGCTACAAAGTTTTGCCTCATCTGCCATAT
HSMAR1_Dn CCCAATTCTTCTCCACGACAACGCCCGACCGCACGTCGCACAACCAATGCTTCAAAAGTTGAACGAATTGGGCTATGAAGTTTTGCCTCATCCGCCATAT
MADE1
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
HSMAR1 TCACCTGACCTCTCGCCAACCGACTACCACTTCTTCAAGCATCTCGACAACTTTTTGCAGGGAAAACGCTTCCACAACCAGCAGGATGCAGAAAATGCTT
HSMAR1_Tb TCACCTGACCTCTCGCCAACCGACTACCACTTCTTCAAGCATCTCGACAACTTTTTGCAGGGAAAACACTTCCACAACCAGCAGGATGCAGAAGATGCTT
HSMAR1_Og TCACCTGACCTCTCGCCAACCGACTACCACTTCTTCAAGCATCTCAACAACTTTTTGCAGGGAAAACGCTTCCACAACCAGCAGGATGCAGAAAATGCTT
HSMAR1_Bt TCACCTGACCTCTCGCCAACCGACTACCACTTCTTCAAGCATCTCGACAACTTTTTGCAGGGAAAATGCTTCCACAACCAGCAGGAGGCAGAAAATGCTT
HSMAR1_Dn TCACCTGACCTCTCGCCAACCGACTACCACTTCTTCAAGCATCTCGACAACTTTTTGCAGGGAAAATGCTTCCACAACCAGCAGGATGCAGAAAATGCTT
MADE1
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
HSMAR1 TCCAAGAGTTCGTCGAATCCCGAAGCACGGATTTTTACGCTACAGGAATAAACAAACTTATTTCTCGTTGGCAAAAATGTGTTGATTGTAATGGTTCCTA
HSMAR1_Tb TCCAAGAGTTCGTCGAATCCCGAAGCATGGATTTTTATGCTACAGGAATAAACAAACTTATTTCTCATTGGCAAAAATGTGTTGATTGTAATGGTTCCTA
HSMAR1_Og TCCAAGAGTTCGTCGAATCCCGAAGCATGGATTTTTATGCTACAGGAAGAAACAAACTTATTTCTCATTGGCAACAATGTGTTGATTGTAATGGTTCCTA
HSMAR1_Bt TCCAAGAGTTCGTCGAATCCCGAAGCATGGATTTTTATGCTACAGGAATAAACAAACTTATTTCTCATTGGCAAAAATGTGTTGATTGTAATGGTTCCTA
HSMAR1_Dn TCCAAGAGTTCGTCGAATCCCGAAGCACAGATTTTTATGCTACAGGAATAAACAAACTTATTTCTCATTGGCAAAAATGTGTTGATTGTAATGGTTCCTA
MADE1
1210 1220 1230 1240 1250 1260 1270 1280
HSMAR1 TTTTGATTAATAAAGATGTGTTTGAGCCTAGTTATAATGATTTAAAATTCACGGTCCAAAACCGCAATTACTTTTGCACCAACCTAA
HSMAR1_Tb TTTTGATTAATAAAGATGTGTTTGAGCCTAGTTATAATGATTTAAAATTCATGGTCCAAAACTACAATTGCTTTTGCACCAACCTAA
HSMAR1_Og TTTTGATTAATAAAGATGTGTTTGAGCCTAGTTATAATGATTTAAAATTCATGGTCCAAAACCACAATTACTTTTGCACCAACCTAA
HSMAR1_Bt TTTTGATTAATAAAGATGTGTTTGAGCCTAGTTATAATGATTTAAAATTCACGGTCCAAAACTGCAATTACTTTTGCACCAACCTAA
HSMAR1_Dn TTTTGATTAATAAAGATGTGTTTGAGCCTAGTTATAATGATTTAAAATTCACGGTCCGAAACCGCAATTCCTTTTGCACCAACCTAA
MADE1 ----------catttactttyaatggcaaaaaccgcaattactttgcaccaacctaa

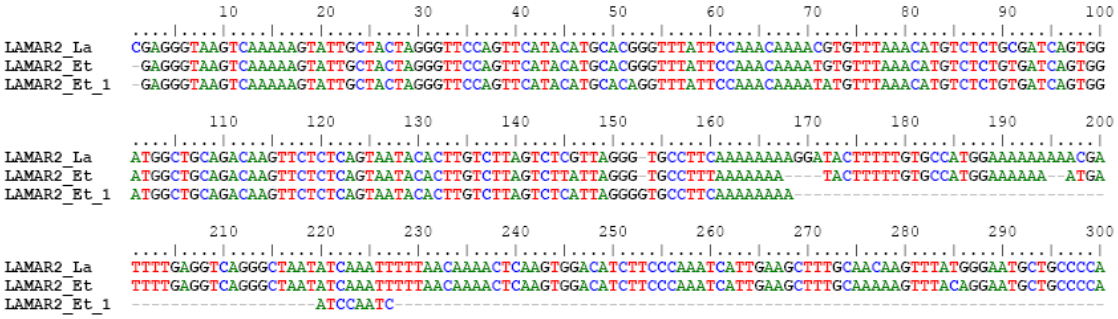
HSMAR2 Tribe

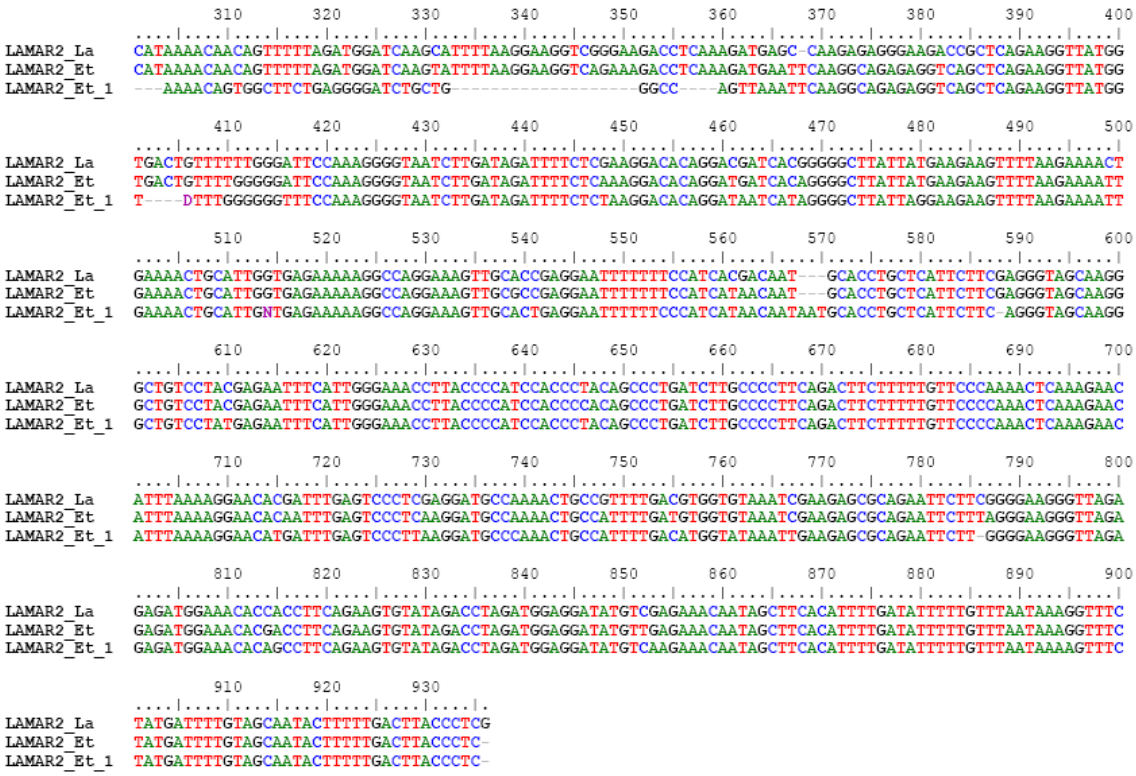



610 620 630 640 650 660 670 680 690 700
HSMAR2 TGGAAATTTTAAACAAGTGGGATCAAGATCCTGAAGCATTTCTTCGAAGAATTGTAACAGGAGATGAAACGTGGCTTTACCAGTACGATCCTGAAGACAA
HSMAR2_Tb TGGAAATTTTAAACAAGTGGGATCAAGATCCTGAAGCATTTCTTCGAAAAATTGTAACAGGAGATGACACATGGCTTTACCAGTATGATCCTGAAGACAA
HSMAR2_Oc TGGAAATTTTAAACAAGTGGGATCAAGATCCTGAAGCATTTCTTCNAAGAATTGTAACAGGAGATGAAACATGGCTTTACCAGTATGATCCTGAAGACAA
HSMAR2_Og TGGAAATTTTAAACAAGTGGGATCAAGATCCTGAAGCATTTCTTCGAAGAATTGTAACAGGAGATGAAACATGGCTTTACCAGTACAATCCTGAAGACAA
HSMAR2_Et TGGAAATTTTAAACAAGTGGGATCAAGATCCTGAAGCATTTCTTGGAAGAATTGTAACAGGAGATGAAACATGGCTTTACCAGTATGATCCTGAAGACAA
HSMAR2_Tb_1
HSMAR2_Op_1
HSMAR2_Oc_1
HSMAR2_Et_1
710 720 730 740 750 760 770 780 790 800
HSMAR2 AGCACAATCAAAGCAATGGCTACCAAGAGGTGGAAGTGGTCCAGTCAAAGCAAAAGCGGACTGGTCAAGAGCAAAGGTCATGGCAACAGTTTTTTGGGAT
HSMAR2_Tb AGCACAATCAAAGCAATGGCTACCAAGAGGTGGAAGTGGTCCAGTCAAAGCAAAAGCGGACTGGTCAAGAGCAAAGGTCATGGCAACAGTTTTTTGGGAT
HSMAR2_Oc AGCACAATCAAAGCAATGGCTACCAAGAGGTGGAAGTGGTCCAGTCAAAGCAAAAGCAGACCGGTCAAGAGCAAAGGTCATGGCAACAGTTTTTTGGGAT
HSMAR2_Og AGCACAATCAAAGCAATGGCTACCAAGAGGTGGAAGTGGTCCAGTCAAAGCAAAAGCAGACCAGTCAAGAGCAAAGGTCATGGCAACAGTTTTTTGGGAT
HSMAR2_Et AGCACAATCAAAGCAATGGCTACCAAGAGGTGGAAGTGGTCCAGTCAAAGCAAAAGCGGACCGGTCAAGAGCAAAGGTCATGGCAACAGTTTTTTGGGAT
HSMAR2_Tb_1
HSMAR2_Op_1
HSMAR2_Oc_1
HSMAR2_Et_1
810 820 830 840 850 860 870 880 890 900
HSMAR2 GCTCAAGGCATTTTGCTTGTTGACTTTCTGGAGGGCCAAAGAACGATAACATCTGCTTATTATGAGAGTGTTTTGAGAAAGTTAGCCAAAGCTTTAGCAG
HSMAR2_Tb GCTCAAGGCATTTTGCTTATTGACTTTCTGGAAGGCCAAAGAATAATAACATCTGCTTATTATGAGAGTGTTTTGAGAAAGTTAGCCGAAGCTTTAGCAG
HSMAR2_Oc GTTCAAGGCATTTTGCTTGTTGACTTTCTGGAGGGCCAAAGAATGATAACATCTGCTTATTATAAGAGTGTTTTGAGAAAGTTAGCCAAAGCTTTAGCAG
HSMAR2_Og GCTCAAGGCATTTTGCTTGTTGACTTTCTGGAGGACCAAAGAATAATAACATCTGCTTATTATGAGAGTGTTTTGAGAAAGTTAGCCAAAGCTTTAGCAG
HSMAR2_Et GCTCAAGGCATTTTGCTTGTTGACTTTCTGGAGGGCCAAAGAATAATAACATCTGCTTATTATGAGAGTGTTTTGAGAAAGTTAGCCAAAGCTTTAGCAG
HSMAR2_Tb_1
HSMAR2_Op_1
HSMAR2_Oc_1
HSMAR2_Et_1
910 920 930 940 950 960 970 980 990 1000
HSMAR2 AAAAACGCCCGGGAAAGCTTCACCAGAGAGTCCTTCTCCACCACGACAATGCTCCTGCTCATTCCTCTCATCAAACAAGGGCAATTTTGCGAGAGTTTCG
HSMAR2_Tb AAAAACGCCCGGGAAAGCTTCACCAGAGAGTCCTTCTTCACCACGACAATGCTCCTGCTCATTCCTCTCATCAAACAAGGGCAATTTTGCGAGAGTTTCG
HSMAR2_Oc AAAAACGCCCAGGAAAGCTTCACCAGAGAGTCCTTCTCCACCATGACAATGCTCCTGCTCATTCCTCTCATCAAACAAGGGCAATTTTGCGAGAGTTTCG
HSMAR2_Og AAAAATGCCCAGGAAAGCTTCACCAGAGAGTCCTTCTCCACCATGACAATGCTCCTGCTCATTCCTCTCATCAAACAAGGGCAATTTTGCGAGAGTTTCG
HSMAR2_Et AAAAACACCCAGGAAAGCTTCACCAGAGAGTCCTTCTCCACCACAACAATGCTCCTGCTCATTCCTCTCATCAAACAAGGGCAATTTTGCGAGAGTTTCG
HSMAR2_Tb_1
HSMAR2_Op_1
HSMAR2_Oc_1
HSMAR2_Et_1
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
HSMAR2 ATGGGAAATCATTAGGCATCCACCTTACAGTCCTGATTTGGCTCCTTCTGACTTCTTTTTGTTTCCTAATCTTAAAAAATCTTTAAAGGGCACCCATTTT
HSMAR2_Tb ATGGGAAATCATTAGGCATCCACCTTACAGTCCTGATTTGGCTCCTTCTNACTTHTTTTTGTTTCCTAATCTTAAAAAATCTTTAAAGGGCACCCATTTT
HSMAR2_Oc ATGGGAAATCATTAGGCATCCACCTTACAGTCCTGATTTGGCTCCTTCTGACTTCTTTTTGTTTCCTAATCTTAAAAAATCTTTAAAGGGCACCCATTTT
HSMAR2_Og ATGGGAAATCATTAGGCATCCACCTTACAGTCCTGATTTGGCTCCTTCTGACTTCTTTTTGTTTCCTAATCTTAAAAAATCTTTAAAGGGCACCCATTTT
HSMAR2_Et ATGGGAAATCATTAGGCATCTACCTTACAGTCCTGATTTGGCTCCTTCTGANTTCTTTTTGTTTCCAATTCTTAAAAAATCTTTAAAGGCCCCCCTTTTT
HSMAR2_Tb_1
HSMAR2_Op_1
HSMAR2_Oc_1
HSMAR2_Et_1
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
HSMAR2 TCTTCAGTTAATAATGTAAAAAAGACTGCATTGACATGGTTAAATTCCCAGGACCCTCAGTTCTTTAGGGATGGACTAAATGGCTGGTATCATCGCTTAC
HSMAR2_Tb TCTTCADTTAATAATGTAAAAAAGACTGCATTGACATGGTTAAATTCCCAGGACCCTCAGTTCTTTAGGGATGGACTAAATGGCTGGTATCATNGCTTAC
HSMAR2_Oc TCTTCAGTTAATAATGTAAAAAAAACTGCATTGACATGGTTAAATTCCCAGGACCCTCAGTTCTTTAGGGATGGACTAAATGGCTGGTATCATCACTTAC
HSMAR2_Og TCTTCAGTTAATAATGTAAAAAAGACTGCATTGACATGGTTAAATTCCCAGGACCCTCAGTTCTTTAGGGATGGACTAAATGGCTGGTATCATCGCTTAC
HSMAR2_Et TCTTCAGTTAATAATGTAAAAAAGACTGCATTGACATGGTTAAATTCCCAGGACCCTCAGTTCTTTAGGGATGGACTAAATGGCTGGTATCATCACTTAC
HSMAR2_Tb_1
HSMAR2_Op_1
HSMAR2_Oc_1
HSMAR2_Et_1

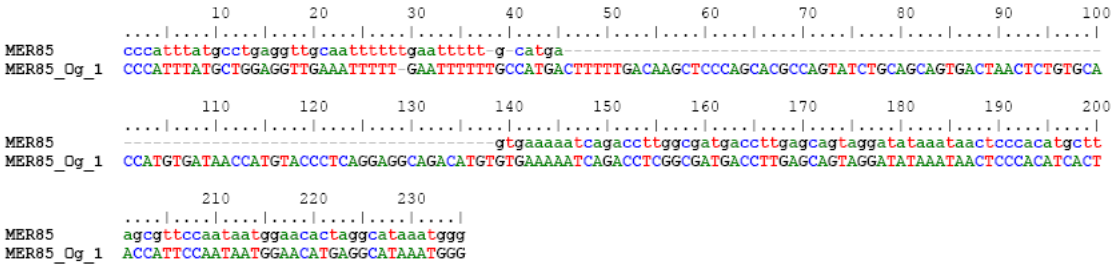


LAMAR2 Tribe

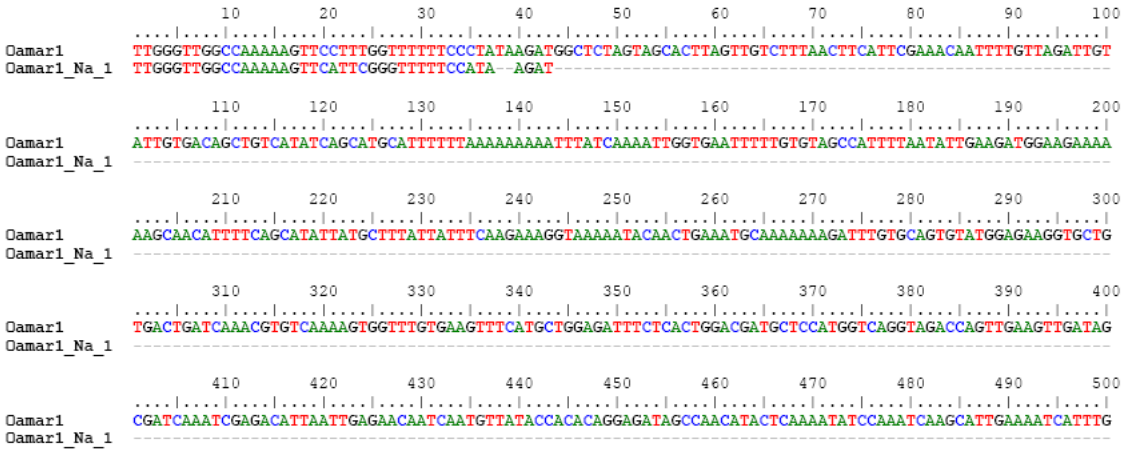


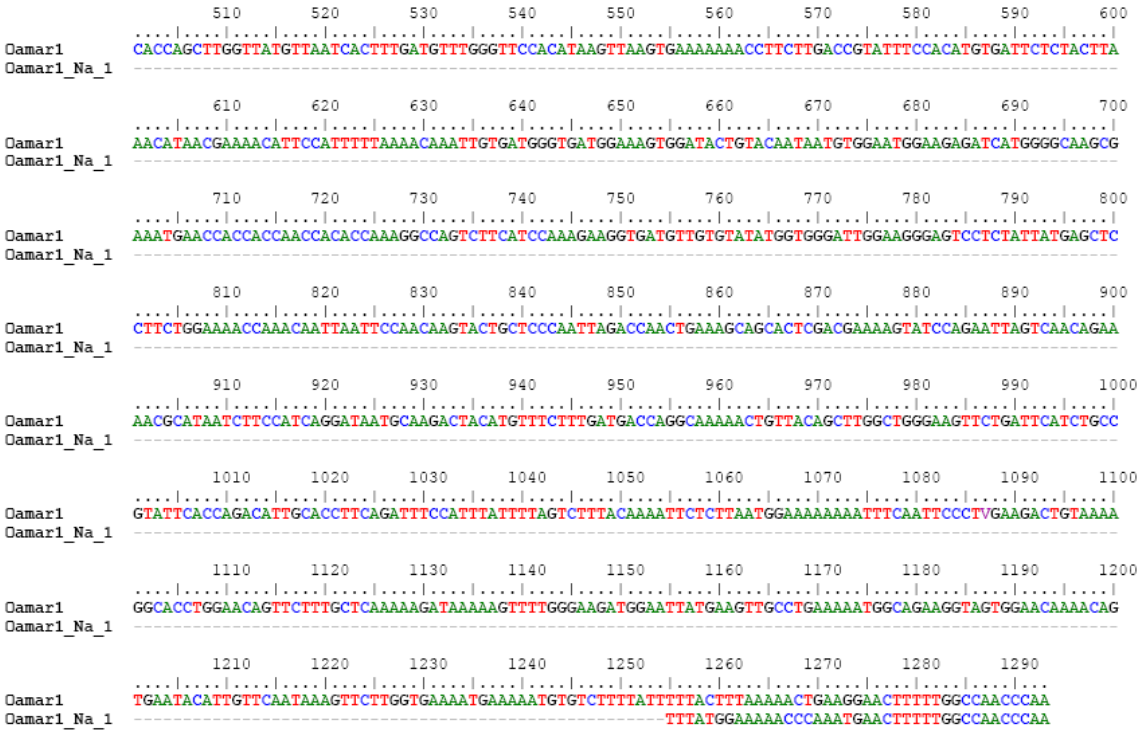

LAMAR2_La
LAMAR2_Et
LAMAR2_Et_1

MER85 Tribe



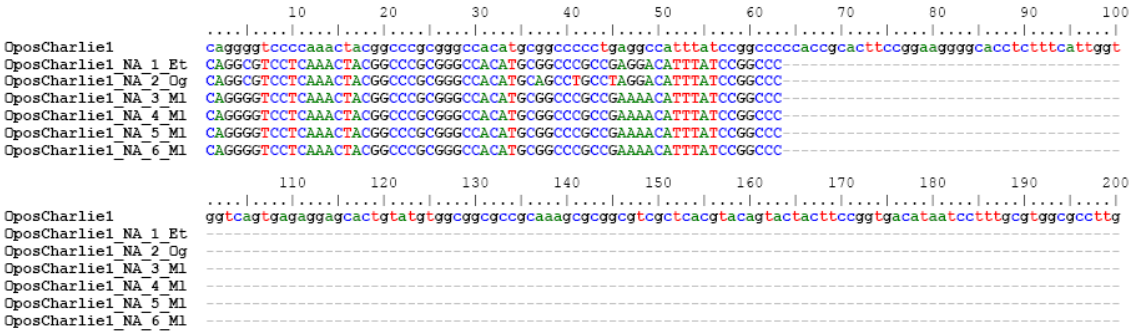
Oamar1 Tribe

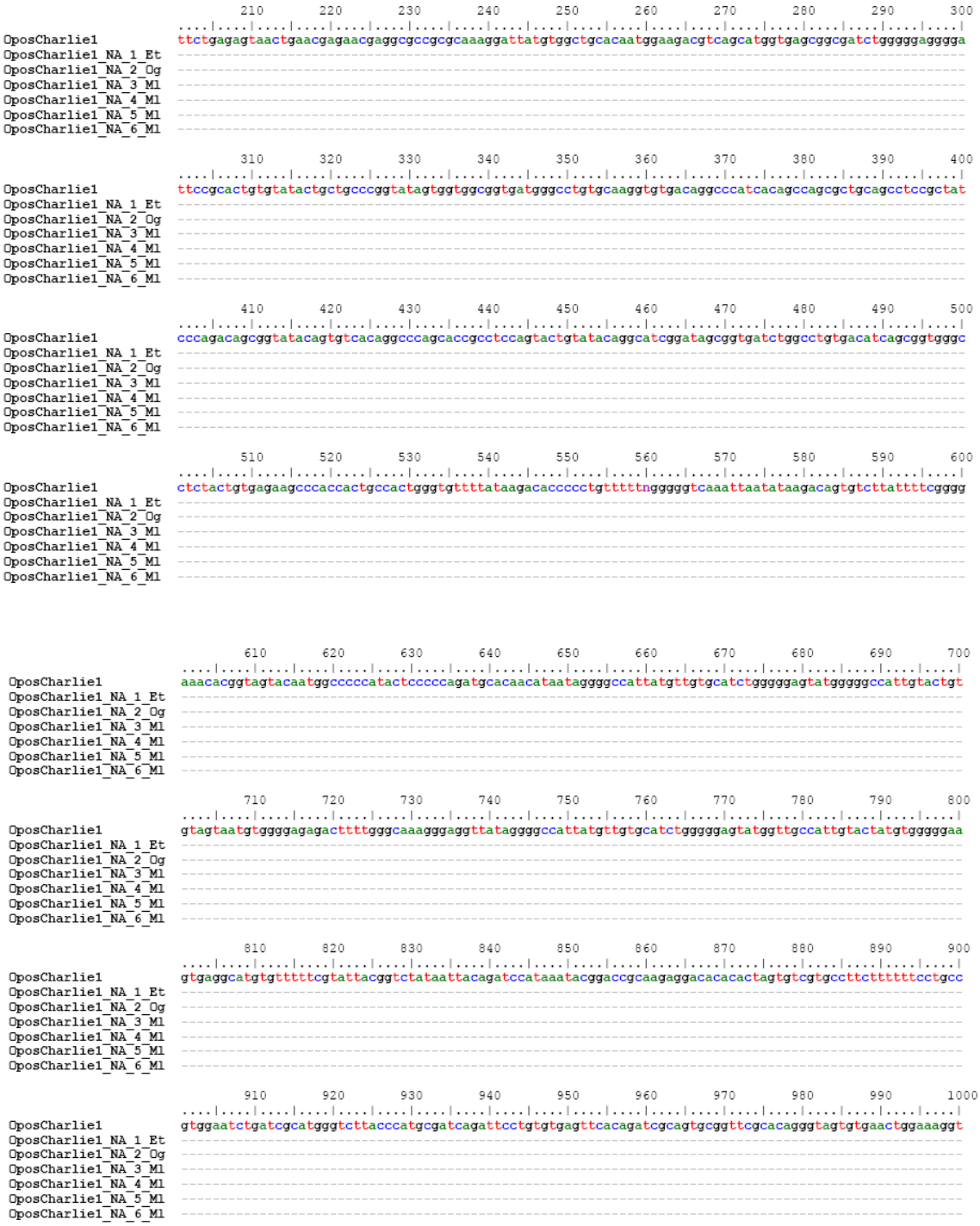


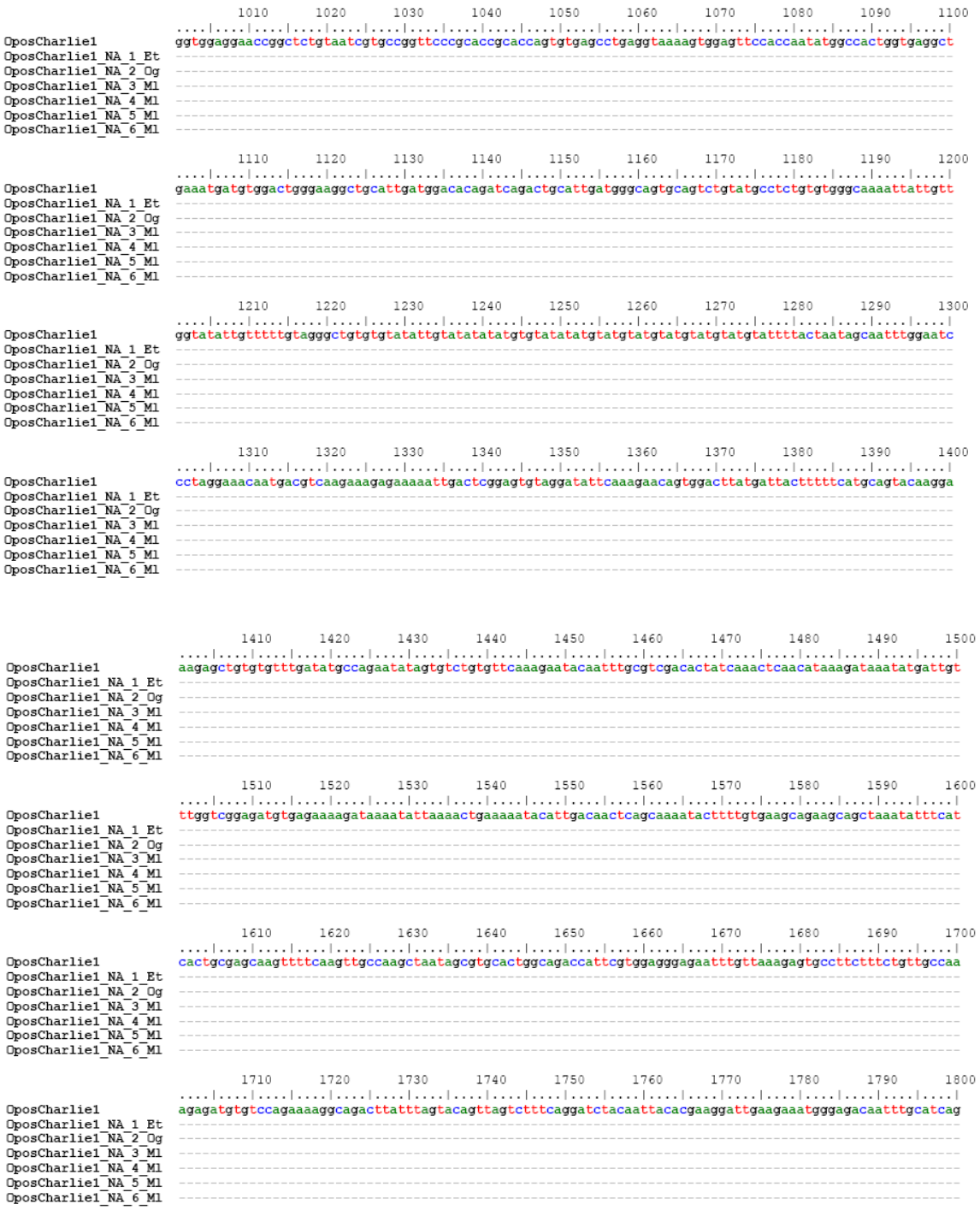

510 520 530 540 550 560 570 580 590 600
Oamar1 CACCAGCTTGGTTATGTTAATCACTTTGATGTTTGGGTTCCACATAAGTTAAGTGAAAAAAACCTTCTTGACCGTATTTCCACATGTGATTCTCTACTTA
Oamar1_Na_1
610 620 630 640 650 660 670 680 690 700
Oamar1 AACATAACGAAAACATTCCATTTTTAAAACAAATTGTGATGGGTGATGGAAAGTGGATACTGTACAATAATGTGGAATGGAAGAGATCATGGGGCAAGCG
Oamar1_Na_1
710 720 730 740 750 760 770 780 790 800
Oamar1 AAATGAACCACCACCAACCACACCAAAGGCCAGTCTTCATCCAAAGAAGGTGATGTTGTGTATATGGTGGGATTGGAAGGGAGTCCTCTATTATGAGCTC
Oamar1_Na_1
810 820 830 840 850 860 870 880 890 900
Oamar1 CTTCTGGAAAACCAAACAATTAATTCCAACAAGTACTGCTCCCAATTAGACCAACTGAAAGCAGCACTCGACGAAAAGTATCCAGAATTAGTCAACAGAA
Oamar1_Na_1
910 920 930 940 950 960 970 980 990 1000
Oamar1 AACGCATAATCTTCCATCAGGATAATGCAAGACTACATGTTTCTTTGATGACCAGGCAAAAACTGTTACAGCTTGGCTGGGAAGTTCTGATTCATCTGCC
Oamar1_Na_1
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
Oamar1 GTATTCACCAGACATTGCACCTTCAGATTTCCATTTATTTTAGTCTTTACAAAATTCTCTTAATGGAAAAAAAATTTCAATTCCCTVGAAGACTGTAAAA
Oamar1_Na_1
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
Oamar1 GGCACCTGGAACAGTTCTTTGCTCAAAAAGATAAAAAGTTTTGGGAAGATGGAATTATGAAGTTGCCTGAAAAATGGCAGAAGGTAGTGGAACAAAACAG
Oamar1_Na_1
1210 1220 1230 1240 1250 1260 1270 1280 1290
Oamar1 TGAATACATTGTTCAATAAAGTTCTTGGTGAAAATGAAAAATGTGTCTTTTATTTTTACTTTAAAAACTGAAGGAACTTTTTGGCCAACCCAA
Oamar1_Na_1 TTTATGGAAAAACCCAAATGAACTTTTTGGCCAACCCAA

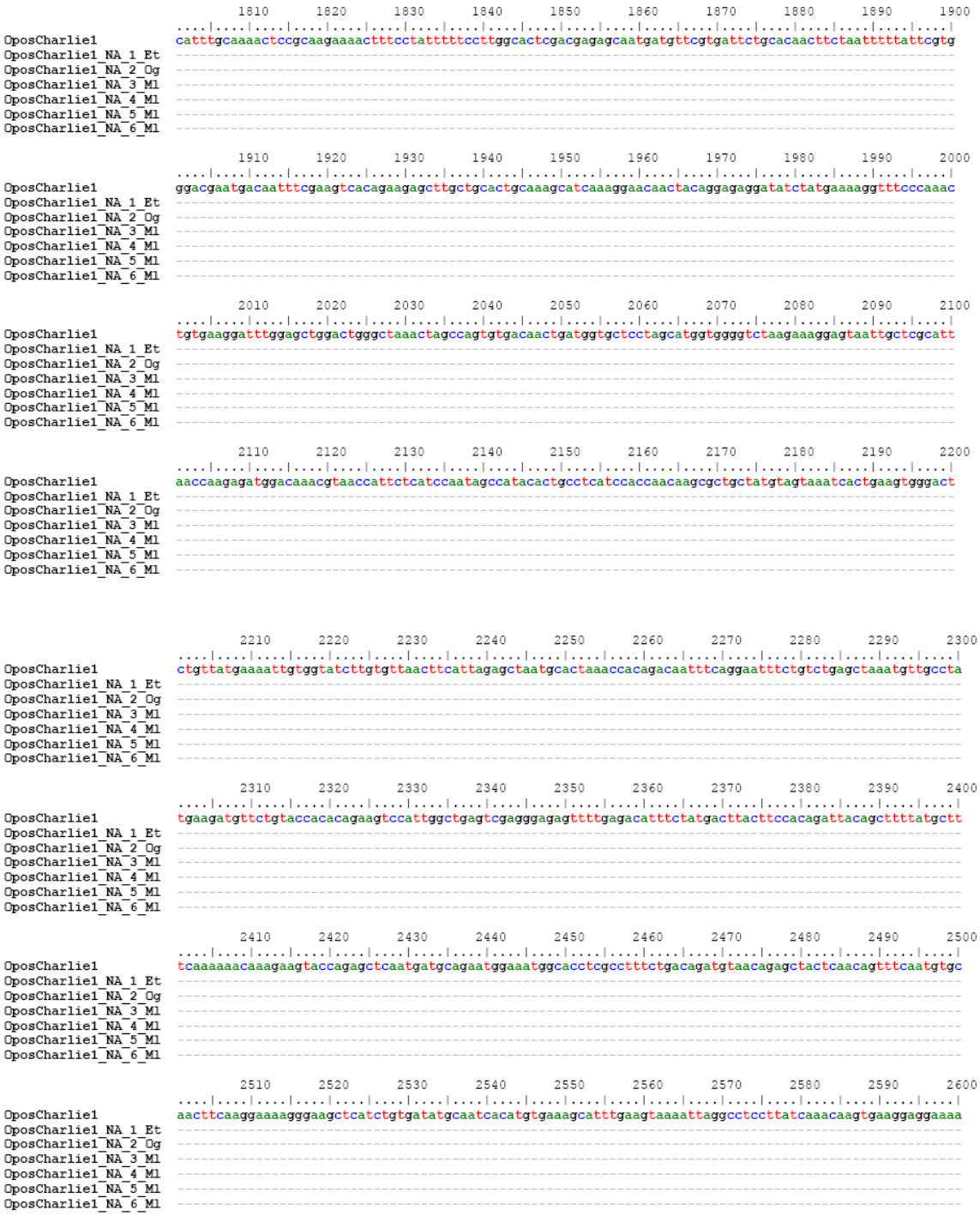
OGMAR2 Tribe

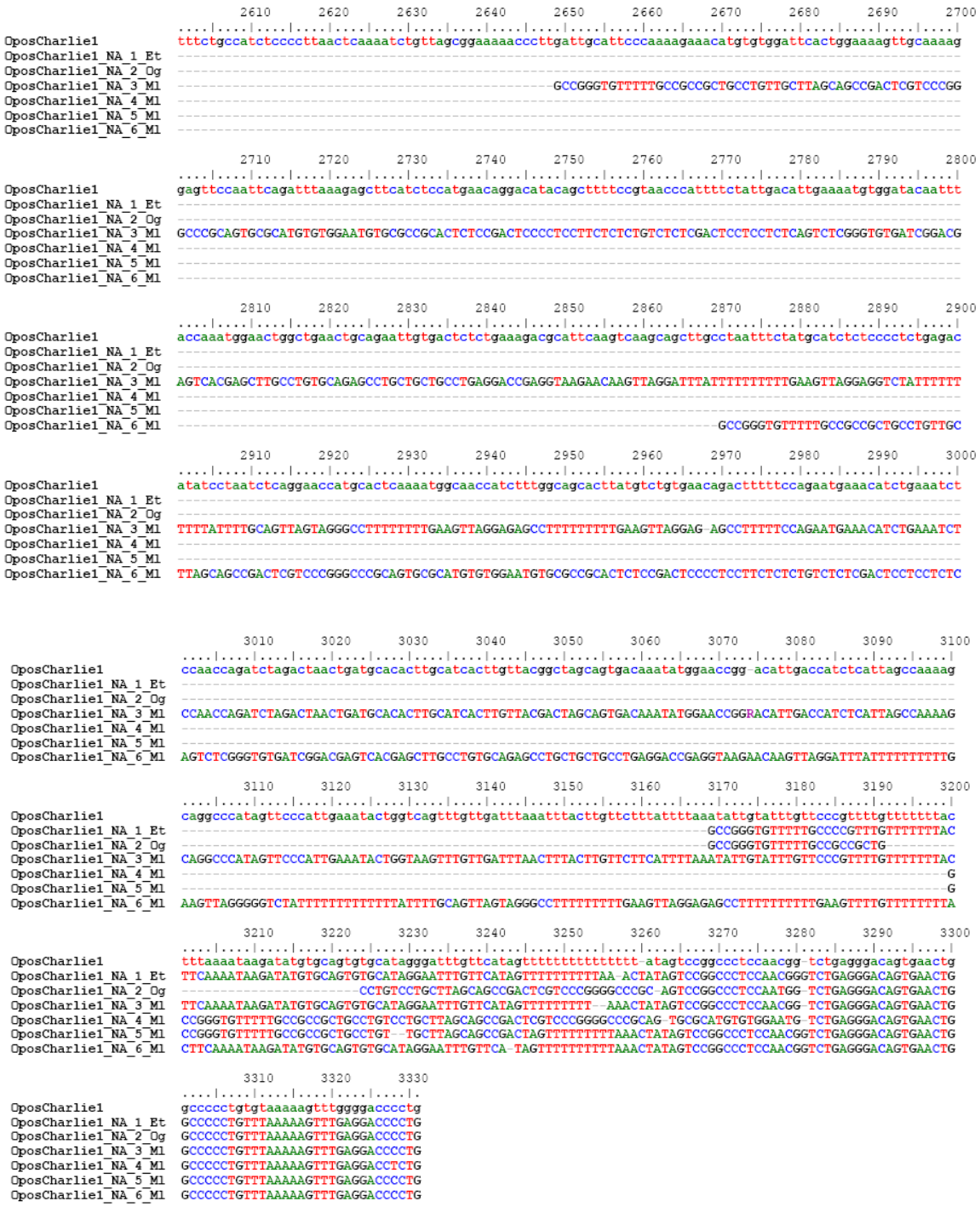
OposCharlie1 Tribe



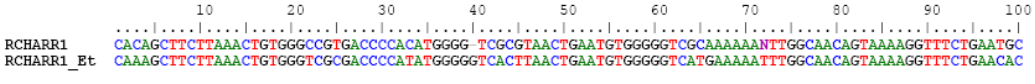

210 220 230 240 250 260 270 280 290 300
OposCharlie1 ttctgagagtaactgaacgagaacgaggcgccgcgcaaaggattatgtggctgcacaatggaagacgtcagcatggtgagcggcgatctgggggagggga
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
310 320 330 340 350 360 370 380 390 400
OposCharlie1 ttccgcactgtgtatactgctgcccggtatagtggtggcggtgatgggcctgtgcaaggtgtgacaggcccatcacagccagcgctgcagcctccgctat
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
410 420 430 440 450 460 470 480 490 500
OposCharlie1 cccagacagcggtatacagtgtcacaggcccagcaccgcctccagtactgtatacaggcatcggatagcggtgatctggcctgtgacatcagcggtgggc
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
510 520 530 540 550 560 570 580 590 600
OposCharlie1 ctctactgtgagaagcccaccactgccactgggtgttttataagacacccccctgtttttngggggtcaaattaatataagacagtgtcttatttttcgggg
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
610 620 630 640 650 660 670 680 690 700
OposCharlie1 aaacacggtagtacaatggccccatactccccagatgcacaacataatagggccattatgttgtgcatctgggggagtatgggggccattgtactgt
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
710 720 730 740 750 760 770 780 790 800
OposCharlie1 gtagtaatgtgggggagagacttttgggcaaagggaggttatagggccattatgttgtgcatctgggggagtatggttgccattgtactatgtgggggaa
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
810 820 830 840 850 860 870 880 890 900
OposCharlie1 gtgaggcatgtgtttttcgtattacggtctataattacagatccataaatacggaccgcaagaggacacacactagtgtcgtgccttctttttttcctgcc
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
910 920 930 940 950 960 970 980 990 1000
OposCharlie1 gtggaatctgatcgcatgggtcttacccatgcgatcagattcctgtgtgagttcacagatcgcagtgcggttcgcacagggtagtgtgaactggaaaggt
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml


1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
OposCharlie1 ggtggaggaaccggctctgtaatcgtgccggttcccgcaccgcaccagtgtgagcctgaggtaaaagtggagttccaccaatatggccactggtgaggct
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
OposCharlie1 gaaatgatgtggactgggaaggctgcattgatggacacagatcagactgcattgatgggcagtgcagtctgtatgcctctgtgtgggcaaaattattgtt
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
OposCharlie1 ggtatattgtttttgtagggctgtgtgtatattgtatatatatgtgtatatatgtatgtatgtatgtatgtatttactaatagcaatttggaatc
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
OposCharlie1 cctaggaaacaatgacgtcaagaaagagaaaaattgactcggagtgtaggatattcaaagaacagtggacttatgattacttttttcatgcagtacaagga
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
OposCharlie1 aagagctgtgtgtttgatatgccagaatatagtgtctgtgttcaaagaatacaatttgcgtcgacactatcaaactcaacataaagataaatatgattgt
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
OposCharlie1 ttggtcggagatgtgagaaaagataaaatattaaaactgaaaaatacattgacaactcagcaaaatacttttgtgaagcagaagcagctaaatatttcat
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
OposCharlie1 cactgcgagcaagttttcaagttgccaagctaatagcgtgcactggcagaccattcgtggagggagaatttgttaaagagtgccttctttctgttgccaa
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
OposCharlie1 agagatgtgtccagaaaaggcagacttatttagtacagttagtctttcaggatctacaattacacgaaggattgaagaaatgggagacaatttgcatcag
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml


1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
OposCharlie1 catttgcaaaactccgcaagaaaacttttcctatttttccttggcactcgacgagagcaatgatgttcgtgattctgcacaacttctaatttttattcgtg
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
OposCharlie1 ggacgaatgacaatttcgaagtcacagaagagcttgctgcactgcaaagcatcaaaggaacaactacaggagaggatatctatgaaaaggtttcccaaac
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
OposCharlie1 tgtgaaggatttggagctggactgggctaaactagccagtgtgacaactgatggtgctcctagcatggtggggtctaagaaaggagtaattgctcgcatt
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
OposCharlie1 aaccaagagatggacaaacgtaaccattctcatccaatagccatacactgcctcatccaccaacaagcgctgctatgtagtaaatcactgaagtgggact
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
OposCharlie1 ctgttatgaaaattgtggtatcttgtgttaacttcattagagctaatgcactaaaccacagacaatttcaggaatttctgtctgagctaaatgttgccta
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
OposCharlie1 tgaagatgttctgtaccacacagaagtccattggctgagtcgagggagagttttgagacatttctatgacttacttccacagattacagcttttatgctt
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
OposCharlie1 tcaaaaaacaaagaagtaccagagctcaatgatgcagaatggaaatggcacctcgcctttctgacagatgtaacagagctactcaacagtttcaatgtgc
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
OposCharlie1 aacttcaaggaaaagggaagctcatctgtgatatgcaatcacatgtgaaagcatttgaagtaaaattaggcctccttatcaaacaagtgaaggaggaaaa
OposCharlie1_NA_1_Et
OposCharlie1_NA_2_Og
OposCharlie1_NA_3_Ml
OposCharlie1_NA_4_Ml
OposCharlie1_NA_5_Ml
OposCharlie1_NA_6_Ml



RCHARR1 Tribe



REFERENCES

Abrusan, G., J. Giordano, and P.E. Warburton. 2008. Analysis of transposon interruptions suggests selection for L1 elements on the X chromosome. *PLoS Genet* **4:** e1000172.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402.

Andersson, J.O. 2005. Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* **62:** 1182-1197.

Auge-Gouillou, C., Y. Bigot, N. Pollet, M.H. Hamelin, M. Meunier-Rotival, and G. Periquet. 1995. Human and other mammalian genomes contain transposons of the mariner family. *FEBS Lett* **368:** 541-546.

Bailey, J.A. and E.E. Eichler. 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet* **7:** 552-564.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W.J. Kent, J.S. Mattick, and D. Haussler. 2004. Ultraconserved elements in the human genome. *Science* **304:** 1321-1325.

Bininda-Emonds, O.R., M. Cardillo, K.E. Jones, R.D. MacPhee, R.M. Beck, R. Grenyer, S.A. Price, R.A. Vos, J.L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* **446:** 507-512.

Cabot, E.L., B. Angeletti, K. Usdin, and A.V. Furano. 1997. Rapid evolution of a young L1 (LINE-1) clade in recently speciated Rattus taxa. *J Mol Evol* **45:** 412-423.

Caceres, M., J.M. Ranz, A. Barbadilla, M. Long, and A. Ruiz. 1999. Generation of a widespread Drosophila inversion by a transposable element. *Science* **285:** 415-418.

Caspi, A. and L. Pachter. 2006. Identification of transposable elements using multiple alignments of related genomes. *Genome Res* **16:** 260-270.

Cheng, Z., M. Ventura, X. She, P. Khaitovich, T. Graves, K. Osoegawa, D. Church, P. DeJong, R.K. Wilson, S. Paabo et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437:** 88-93.

Clark, J.B., W.P. Maddison, and M.G. Kidwell. 1994. Phylogenetic analysis supports horizontal transfer of P transposable elements. *Mol Biol Evol* **11:** 40-50.

Clark, J.B. and M.G. Kidwell. 1997. A phylogenetic perspective on P transposable element evolution in Drosophila. *Proc Natl Acad Sci U S A* **94:** 11428-11433.

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

Cordaux, R., S. Udit, M.A. Batzer, and C. Feschotte. 2006b. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci U S A* **103:** 8101-8106.

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

Crawford, A.J. 2003. Relative rates of nucleotide substitution in frogs. *J Mol Evol* **57:** 636-641.

D'Anjou, H., C. Chabot, and P. Chartrand. 2004. Preferential accessibility to specific genomic loci for the repair of double-strand breaks in human cells. *Nucleic Acids Res* **32:** 6136-6143.

Daniels, S.B., K.R. Peterson, L.D. Strausbaugh, M.G. Kidwell, and A. Chovnick. 1990. Evidence for horizontal transmission of the P transposable element between Drosophila species. *Genetics* **124:** 339-355.

de Boer, J.G., R. Yazawa, W.S. Davidson, and B.F. Koop. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* **8:** 422.
Decottignies, A. 2005. Capture of extranuclear DNA at fission yeast double-strand breaks. *Genetics* **171:** 1535-1548.
Deininger, P.L., J.V. Moran, M.A. Batzer, and H.H. Kazazian, Jr. 2003. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* **13:** 651-658.
Delsuc, F., S.F. Vizcaino, and E.J. Douzery. 2004. Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evol Biol* **4:** 11.
Dematteï, M.V., C. Auge-Gouillou, N. Pollet, M.H. Hamelin, M. Meunier-Rotival, and Y. Bigot. 2000. Features of the mammal mar1 transposons in the human, sheep, cow, and mouse genomes and implications for their evolution. *Mamm Genome* **11:** 1111-1116.
Diao, X., M. Freeling, and D. Lisch. 2006. Horizontal transfer of a plant transposon. *PLoS Biol* **4:** e5.
Eichler, E.E. and D. Sankoff. 2003. Structural dynamics of eukaryotic chromosome evolution. *Science* **301:** 793-797.
Eickbush, T.H. and H.S. Malik. 2002. Origins and evolution of retrotransposons. In *Mobile DNA II* (ed. N.L.C.e. al.), pp. 1111-1144. ASM Press, Washington, DC.
Emerman, M. 2006. How TRIM5alpha defends against retroviral invasions. *Proc Natl Acad Sci U S A* **103:** 5249-5250.
Evertts, A.G., C. Plymire, N.L. Craig, and H.L. Levin. 2007. The hermes transposon of Musca domestica is an efficient tool for the mutagenesis of Schizosaccharomyces pombe. *Genetics* **177:** 2519-2523.
Feschotte, C., X. Zhang, and S.R. Wessler. 2002. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In *Mobile DNA II* (eds. N.L. Craig R. Craigie M. Gellert, and A. Lambowitz). American Society of Microbiology Press, Washington, DC.
Feschotte, C. 2004. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. *Mol Biol Evol* **21:** 1769-1780.
Feschotte, C. and E.J. Pritham. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41:** 331-368.
Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9:** 397-405.
Feuk, L., C.R. Marshall, R.F. Wintle, and S.W. Scherer. 2006. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* **15 Spec No 1:** R57-66.
Fiston-Lavier, A.S., D. Anxolabehere, and H. Quesneville. 2007. A model of segmental duplication formation in Drosophila melanogaster. *Genome Res* **17:** 1458-1470.
Fraser, M.J., G.E. Smith, and M.D. Summers. 1983. Acquisition of Host Cell DNA Sequences by Baculoviruses: Relationship Between Host DNA Insertions and FP Mutants of Autographa californica and Galleria mellonella Nuclear Polyhedrosis Viruses. *J Virol* **47:** 287-300.
Friesen, P.D. and M.S. Nissen. 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. *Mol Cell Biol* **10:** 3067-3077.

Furano, A.V. 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog Nucleic Acid Res Mol Biol* **64:** 255-294.
Gentles, A.J., M.J. Wakefield, O. Kohany, W. Gu, M.A. Batzer, D.D. Pollock, and J. Jurka. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica. *Genome Res* **17:** 992-1004.
Gifford, R. and M. Tristem. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26:** 291-315.
Giordano, J., Y. Ge, Y. Gelfand, G. Abrusan, G. Benson, and P.E. Warburton. 2007. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comput Biol* **3:** e137.
Gladyshev, E.A., M. Meselson, and I.R. Arkhipova. 2008. Massive horizontal gene transfer in Bdelloid rotifers. *Science* **320:** In Press.
Gonzalez, P. and H.A. Lessios. 1999. Evolution of sea urchin retroviral-like (SURL) elements: evidence from 40 echinoid species. *Mol Biol Evol* **16:** 938-952.
Goodman, M. 1985. Rates of molecular evolution: the hominoid slowdown. *Bioessays* **3:** 9-14.
Goodman, M. 1999. The genomic record of Humankind's evolutionary roots. *Am J Hum Genet* **64:** 31-39.
Gorbunova, V. and A.A. Levy. 1997. Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res* **25:** 4650-4657.
Gray, Y.H. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet* **16:** 461-468.
Haber, J.E. 2000. Partners and pathways: repairing a double-strand break. *Trends in Genetics* **16:** 259-264.
Hagemann, S. and W. Pinsker. 2001. Drosophila P transposons in the human genome? *Mol Biol Evol* **18:** 1979-1982.
Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41:** 95-98.
Han, K., M.K. Konkel, J. Xing, H. Wang, J. Lee, T.J. Meyer, C.T. Huang, E. Sandifer, K. Hebert, E.W. Barnes et al. 2007. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. *Science* **316:** 238-240.
Haring, E., S. Hagemann, and W. Pinsker. 2000. Ancient and recent horizontal invasions of drosophilids by P elements. *J Mol Evol* **51:** 577-586.
Hartl, D.L., A.R. Lohe, and E.R. Lozovskaya. 1997. Modern thoughts on an ancyent marinere: function, evolution, regulation. *Annu Rev Genet* **31:** 337-358.
Hazkani-Covo, E. and D. Graur. 2007. A comparative analysis of numt evolution in human and chimpanzee. *Mol Biol Evol* **24:** 13-18.
Hedges, S.B. and S. Kumar. 2004. Precision of molecular time estimates. *Trends Genet* **20:** 242-247.
Hickey, D.A. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101:** 519-531.
Hotopp, J.C., M.E. Clark, D.C. Oliveira, J.M. Foster, P. Fischer, M.C. Torres, J.D. Giebel, N. Kumar, N. Ishmael, S. Wang et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317:** 1753-1756.
Hugall, A.F., R. Foster, and M.S. Lee. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst Biol* **56:** 543-563.
Inoue, K. and J.R. Lupski. 2002. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* **3:** 199-242.

Janecka, J.E., W. Miller, T.H. Pringle, F. Wiens, A. Zitzmann, K.M. Helgen, M.S. Springer, and W.J. Murphy. 2007. Molecular and genomic data identify the closest living relative of primates. *Science* **318:** 792-794.

Jehle, J.A., A. Nickel, J.M. Vlak, and H. Backhaus. 1998. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into Cydia pomonella granulovirus. *J Mol Evol* **46:** 215-224.

Jordan, I.K. and J.F. McDonald. 1998. Evolution of the copia retrotransposon in the Drosophila melanogaster species subgroup. *Mol Biol Evol* **15:** 1160-1171.

Jordan, I.K., L.V. Matyunina, and J.F. McDonald. 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc Natl Acad Sci U S A* **96:** 12621-12625.

Jurka, J., V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, and J. Walichiewicz. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110:** 462-467.

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

Kapitonov, V.V. and J. Jurka. 2004. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. *DNA Cell Biol* **23:** 311-324.

Kapitonov, V.V. and J. Jurka. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* **3:** e181.

Kawakami, K. 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol* **8 Suppl 1:** S7.

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

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

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

Khan, H., A. Smit, and S. Boissinot. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* **16:** 78-87.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16:** 111-120.

Koga, A., A. Shimada, A. Shima, M. Sakaizumi, H. Tachida, and H. Hori. 2000. Evidence for recent invasion of the medaka fish genome by the Tol2 transposable element. *Genetics* **155:** 273-281.

Kohler, M. and S. Moya-Sola. 1999. A finding of oligocene primates on the European continent. *Proc Natl Acad Sci U S A* **96:** 14664-14667.

Kordis, D. and F. Gubensek. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc Natl Acad Sci U S A* **95:** 10704-10709.

Kriegs, J.O., G. Churakov, M. Kiefmann, U. Jordan, J. Brosius, and J. Schmitz. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol* **4:** e91.

Kulpa, D.A. and J.V. Moran. 2006. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol* **13:** 655-660.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5:** 150-163.

Lander, E.S. L.M. Linton B. Birren C. Nusbaum M.C. Zody J. Baldwin K. Devon K. Dewar M. Doyle W. FitzHugh et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4:** 203-221.
Liang, F., M. Han, P.J. Romanienko, and M. Jasin. 1998. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc Natl Acad Sci U S A* **95:** 5172-5177.
Lim, J.K. and M.J. Simmons. 1994. Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. *Bioessays* **16:** 269-275.
Lin, Y. and A.S. Waldman. 2001. Capture of DNA sequences at double-strand breaks in mammalian chromosomes. *Genetics* **158:** 1665-1674.
Lindblad-Toh, K. C.M. Wade T.S. Mikkelsen E.K. Karlsson D.B. Jaffe M. Kamal M. Clamp J.L. Chang E.J. Kulbokas, 3rd M.C. Zody et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438:** 803-819.
Loftus, B., I. Anderson, R. Davies, U.C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R.P. Hirt, B.J. Mann et al. 2005. The genome of the protist parasite Entamoeba histolytica. *Nature* **433:** 865-868.
Loreto, E.L., C.M. Carareto, and P. Capy. 2008. Revisiting horizontal transfer of transposable elements in Drosophila. *Heredity* **100:** 545-554.
Luan, D.D., M.H. Korman, J.L. Jakubczak, and T.H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72:** 595-605.
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:** 3882-3891.
Malik, H.S., W.D. Burke, and T.H. Eickbush. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16:** 793-805.
Malik, H.S., S. Henikoff, and T.H. Eickbush. 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* **10:** 1307-1318.
Maruyama, K. and D.L. Hartl. 1991. Evidence for interspecific transfer of the transposable element mariner between Drosophila and Zaprionus. *J Mol Evol* **33:** 514-524.
McFadden, G. 2005. Poxvirus tropism. *Nat Rev Microbiol* **3:** 201-213.
Medstrand, P., L.N. van de Lagemaat, and D.L. Mager. 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* **12:** 1483-1495.
Meng, J. and M.C. McKenna. 1998. Faunal turnovers of Palaeogene mammals from the Mongolian Plateau. *Nature (London)* **394:** 364-367.
Messer, P.W. and P.F. Arndt. 2007. The majority of recent short DNA insertions in the human genome are tandem duplications. *Mol Biol Evol* **24:** 1190-1197.
Moore, J.K. and J.E. Haber. 1996. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383:** 644-646.
Morgan, G.T. 1995. Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J Mol Biol* **254:** 1-5.
Morrison, H.G., A.G. McArthur, F.D. Gillin, S.B. Aley, R.D. Adam, G.J. Olsen, A.A. Best, W.Z. Cande, F. Chen, M.J. Cipriano et al. 2007. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* **317:** 1921-1926.
Murphy, W.J., P.A. Pevzner, and S.J. O'Brien. 2004. Mammalian phylogenomics comes of age. *Trends Genet* **20:** 631-639.
Murphy, W.J., T.H. Pringle, T.A. Crider, M.S. Springer, and W. Miller. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res* **17:** 413-421.

Nassif, N., J. Penney, S. Pal, W.R. Engels, and G.B. Gloor. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol Cell Biol* **14:** 1613-1625.
Newman, J.C., A.D. Bailey, H.Y. Fan, T. Pavelitz, and A.M. Weiner. 2008. An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. *PLoS Genet* **4:** e1000031.
Ochman, H., J.G. Lawrence, and E.A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299-304.
Oosumi, T., W.R. Belknap, and B. Garlick. 1995. Mariner transposons in humans. *Nature* **378:** 672.
Orgel, L.E. and F.H. Crick. 1980. Selfish DNA: the ultimate parasite. *Nature* **284:** 604-607.
Pace, J.K., 2nd and C. Feschotte. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* **17:** 422-432.
Pace, J.K., 2nd, C. Gilbert, M.S. Clark, and C. Feschotte. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci U S A* **In press**.
Pavlicek, A. and J. Jurka. 2006. Positive selection on the nonhomologous end-joining factor Cernunnos-XLF in the human lineage. *Biol Direct* **1:** 15.
Phillips, J.W. and W.F. Morgan. 1994. Illegitimate recombination induced by DNA double-strand breaks in a mammalian chromosome. *Mol Cell Biol* **14:** 5794-5803.
Piskurek, O. and N. Okada. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc Natl Acad Sci U S A* **104:** 12046-12051.
Plasterk, R.H. 1991. The origin of footprints of the Tc1 transposon of Caenorhabditis elegans. *EMBO J* **10:** 1919-1925.
Poux, C., O. Madsen, E. Marquard, D.R. Vieites, W.W. de Jong, and M. Vences. 2005. Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Syst Biol* **54:** 719-730.
Poux, C., P. Chevret, D. Huchon, W.W. de Jong, and E.J. Douzery. 2006. Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst Biol* **55:** 228-244.
Price, A.L., E. Eskin, and P.A. Pevzner. 2004. Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res* **14:** 2245-2252.
Pritham, E.J. and C. Feschotte. 2007. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. *Proc Natl Acad Sci U S A* **104:** 1895-1900.
Ray, D.A., H.J. Pagan, M.L. Thompson, and R.D. Stevens. 2007. Bats with hATs: evidence for recent DNA transposon activity in genus Myotis. *Mol Biol Evol* **24:** 632-639.
Ray, D.A., C. Feschotte, H.J. Pagan, J.D. Smith, E.J. Pritham, P. Arensburger, P.W. Atkinson, and N.L. Craig. 2008. Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. *Genome Res* **18:** 717-728.
Ricchetti, M., C. Fairhead, and B. Dujon. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402:** 96-100.
Ricchetti, M., F. Tekaia, and B. Dujon. 2004. Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol* **2:** E273.
Richardson, C. and M. Jasin. 2000a. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405:** 697-700.
Richardson, C. and M. Jasin. 2000b. Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells. *Mol Cell Biol* **20:** 9068-9075.

Robertson, H.M. and E.G. MacLeod. 1993. Five major subfamilies of mariner transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol Biol* **2:** 125-139.
Robertson, H.M. and D.J. Lampe. 1995. Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. *Mol Biol Evol* **12:** 850-862.
Robertson, H.M. 1996. Members of the pogo superfamily of DNA-mediated transposons in the human genome. *Mol Gen Genet* **252:** 761-766.
Robertson, H.M. and R. Martos. 1997. Molecular evolution of the second ancient human mariner transposon, Hsmar2, illustrates patterns of neutral evolution in the human genome lineage. *Gene* **205:** 219-228.
Robertson, H.M. and K.L. Zumpano. 1997. Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. *Gene* **205:** 203-217.
Robertson, H.M. 2002. Evolution of DNA Transposons in Eukaryotes. In *Mobile DNA II* (ed. N.L. Craig). ASM Press.
Salem, A.H., D.A. Ray, D.J. Hedges, J. Jurka, and M.A. Batzer. 2005. Analysis of the human Alu Ye lineage. *BMC Evol Biol* **5:** 18.
Salomon, S. and H. Puchta. 1998. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* **17:** 6086-6095.
Sarkar, A., C. Sim, Y.S. Hong, J.R. Hogan, M.J. Fraser, H.M. Robertson, and F.H. Collins. 2003. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. *Mol Genet Genomics* **270:** 173-180.
Sawyer, S.L. and H.S. Malik. 2006. Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis. *Proc Natl Acad Sci U S A* **103:** 17614-17619.
Saxton, J.A. and S.L. Martin. 1998. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *J Mol Biol* **280:** 611-622.
Silva, J.C. and M.G. Kidwell. 2000. Horizontal transfer and selection in the evolution of P elements. *Mol Biol Evol* **17:** 1542-1557.
Silva, J.C., E.L. Loreto, and J.B. Clark. 2004. Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* **6:** 57-71.
Simmons, G.M. 1992. Horizontal transfer of hobo transposable elements within the Drosophila melanogaster species complex: evidence from DNA sequencing. *Mol Biol Evol* **9:** 1050-1060.
Smit, A., R. Hubley, and P. Green. 1996-2004. RepeatMasker Open-3.0.
Smit, A.F. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res* **21:** 1863-1872.
Smit, A.F., G. Toth, A.D. Riggs, and J. Jurka. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* **246:** 401-417.
Smit, A.F. and A.D. Riggs. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci U S A* **93:** 1443-1448.
Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* **9:** 657-663.
Song, S.U., T. Gerasimova, M. Kurkulos, J.D. Boeke, and V.G. Corces. 1994. An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev* **8:** 2046-2057.

Springer, M.S., W.J. Murphy, E. Eizirik, and S.J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A* **100:** 1056-1061.
Steiper, M.E. and N.M. Young. 2006. Primate molecular divergence dates. *Mol Phylogenet Evol* **41:** 384-394.
Syvanen, M. and C. Kado. 2002. *Horizontal Gene Transfer*. Academic Press, London.
Szak, S.T., O.K. Pickeral, W. Makalowski, M.S. Boguski, D. Landsman, and J.D. Boeke. 2002. Molecular archeology of L1 insertions in the human genome. *Genome Biol* **3:** research0052.
Tarlinton, R.E., J. Meers, and P.R. Young. 2006. Retroviral invasion of the koala genome. *Nature* **442:** 79-81.
Tavare, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. In *Lectures in mathematics in the life sciences*, pp. 57-86. American Mathematical Society, Providence, RI.
Teng, S.-C., B. Kim, and A. Gabriel. 1996. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* **383:** 641-644.
Thomas, E.E., N. Srebro, J. Sebat, N. Navin, J. Healy, B. Mishra, and M. Wigler. 2004. Distribution of short paired duplications in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 10349-10354.
Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876-4882.
Varga, T. and P.D. Aplan. 2005. Chromosomal aberrations induced by double strand DNA breaks. *DNA Repair (Amst)* **4:** 1038-1046.
Waterston, R.H. K. Lindblad-Toh E. Birney J. Rogers J.F. Abril P. Agarwal R. Agarwala R. Ainscough M. Alexandersson P. An et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520-562.
Wei, W., N. Gilbert, S.L. Ooi, J.F. Lawler, E.M. Ostertag, H.H. Kazazian, J.D. Boeke, and J.V. Moran. 2001. Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* **21:** 1429-1439.
Weil, C.F. and R. Kunze. 2000. Transposition of maize Ac/Ds transposable elements in the yeast Saccharomyces cerevisiae. *Nat Genet* **26:** 187-190.
Whitaker Jr., J. and W.J. Hamilton. 1998. *Mammals of the Eastern United States*. Cornell University Press, Ithaca.
Xing, J., A.H. Salem, D.J. Hedges, G.E. Kilroy, W.S. Watkins, J.E. Schienman, C.B. Stewart, J. Jurka, L.B. Jorde, and M.A. Batzer. 2003. Comprehensive analysis of two Alu Yd subfamilies. *J Mol Evol* **57 Suppl 1:** S76-89.
Xing, J., D.J. Hedges, K. Han, H. Wang, R. Cordaux, and M.A. Batzer. 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. *J Mol Biol* **344:** 675-682.
Yang, S., A.F. Smit, S. Schwartz, F. Chiaromonte, K.M. Roskin, D. Haussler, W. Miller, and R.C. Hardison. 2004. Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. *Genome Res* **14:** 517-527.
Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13:** 555-556.
Yi, S., D.L. Ellsworth, and W.H. Li. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. *Mol Biol Evol* **19:** 2191-2198.

Yoder, A.D., M.M. Burns, S. Zehr, T. Delefosse, G. Veron, S.M. Goodman, and J.J. Flynn. 2003. Single origin of Malagasy Carnivora from an African ancestor. *Nature* **421:** 734-737.

Yohn, C.T., Z. Jiang, S.D. McGrath, K.E. Hayden, P. Khaitovich, M.E. Johnson, M.Y. Eichler, J.D. McPherson, S. Zhao, S. Paabo et al. 2005. Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol* **3:** e110.

Yoshiyama, M., Z. Tu, Y. Kainoh, H. Honda, T. Shono, and K. Kimura. 2001. Possible horizontal transfer of a transposable element from host to parasitoid. *Mol Biol Evol* **18:** 1952-1958.

Yu, X. and A. Gabriel. 1999. Patching broken chromosomes with extranuclear cellular DNA. *Mol Cell* **4:** 873-881.

Zennou, V. and P.D. Bieniasz. 2006. Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates. *Virology* **349:** 31-40.

Zhang, J. and T. Peterson. 2004. Transposition of reversed Ac element ends generates chromosome rearrangements in maize. *Genetics* **167:** 1929-1937.

Zhang, X., N. Jiang, C. Feschotte, and S.R. Wessler. 2004. PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. *Genetics* **166:** 971-986.

Zupunski, V., F. Gubensek, and D. Kordis. 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol Biol Evol* **18:** 1849-1863.

BIOGRAPHICAL INFORMATION

John Kelly Pace II was born July 29, 1974. He graduated from the Texas Academy of Mathematics and Science at the University of North Texas with a high school diploma in 1992. He subsequently attended Dallas Baptist University, graduating with a Bachelor of Science in Biology degree in December of 1993. John was married on January 1, 1994, to Deedra Burney. He has 3 daughters, Jana, Julie and Jill.

During graduate school, John's research focused on bioinformatic analysis of the evolution of DNA transposons in mammalian species. After graduate school, he will work as a post-doctoral researcher in the lab of Dr. Sara Sawyer at the University of Texas at Austin, where he will be studying genetic conflict between human and viral genes. John's ultimate career goal is to be a full professor at a major university.