

**LINE DRIVE: L1 ELEMENT ORTHOLOGOUS LOCI**

A Thesis

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

in

The Department of Biological Sciences

By  
Huei Jin Ho  
B.S., Louisiana State University, 2002  
December, 2004

## **ACKNOWLEDGEMENTS**

I would like to thank my advisor Dr. Mark Batzer for his patience, generous financial assistance and supervision that made the completion of my thesis possible. I will always be in debt to him for his many valuable pearls of advices, which I will take with me in my future undertakings. I would also like to extend my gratitude to my thesis committee members, Dr. David Donze and Dr. Mohamed Noor, and all the members of the Batzer Lab, who have been mentors to me in many ways. A special thanks goes to Dr. David Ray, who spared some time from his busy schedule to guide me in my research and thesis preparation. I would also like to thank my family and friends who have always been encouraging and praying for me during the challenging times.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	ii
ABSTRACT.....	vi
1. INTRODUCTION .....	1
1.1. Biology of Retrotransposons.....	1
1.2. The Effects of LINEs on the Human Genome.....	5
1.3. Retrotransposons as Phylogenetic Tools .....	6
1.4. Goal of Study .....	8
2. MATERIALS AND METHODS.....	9
2.1. DNA Samples .....	9
2.2. PCR and Gel Electrophoresis.....	9
2.3. Cloning and Sequencing .....	11
3. RESULTS .....	13
4. DISCUSSION.....	18
REFERENCES .....	23
VITA.....	28

## LIST OF TABLES

<b>2.1.</b> List of DNA sources for all species studied.....	9
<b>2.2.</b> Redesigned preTa L1 oligonucleotide primers .....	10
<b>3.1.</b> Total orthologous loci amplified in non-human primates.....	14
<b>3.2.</b> Simple sequence insertions and genomic deletions .....	14
<b>3.3.</b> L1-mediated deletions of the human genomic sequence .....	16
<b>3.4.</b> Parallel mobile element insertions in non-human primates .....	16

## LIST OF FIGURES

<b>1.1.</b> Structure of a LINE.....	2
<b>1.2.</b> LINEs replicate by a process called Target Primed Reverse Transcription .....	3
<b>3.1.</b> Anomalous events in the L1AD207 locus .....	15
<b>3.2.</b> Parallel insertion of an <i>AluY</i> in the L1AD273 green monkey ortholog .....	17

## ABSTRACT

The L1Hs preTa subfamily is one of the youngest L1 families. It originated after the divergence of human and chimpanzee about 2.34 mya, and therefore is only found in the human genome. Some elements were inserted so recently that they are not fixed in the population. Thirty three of the 254 L1Hs preTa elements are polymorphic for the absence/presence of the insertion, making them useful markers for studying phylogenetics and human population genetics. However, the problem of homoplasy can diminish the value of using L1 elements as phylogenetic and population genetic markers. Examination of the L1Hs preTa orthologous insertion sites in a range of non-human primates revealed an assortment of events that altered the size of the pre-integration or “empty” sites. Only two cases of parallel mobile element insertions into the same pre-integration sites were discovered, one involves an *AluY* in green monkey and the other a L1PA8 element in owl monkey. However, both elements were clearly distinguishable from their human counterparts. No preTa L1 element gene conversion events were observed in any of the loci analyzed. Therefore, we conclude that L1 elements are homoplasy-free genetic characters.

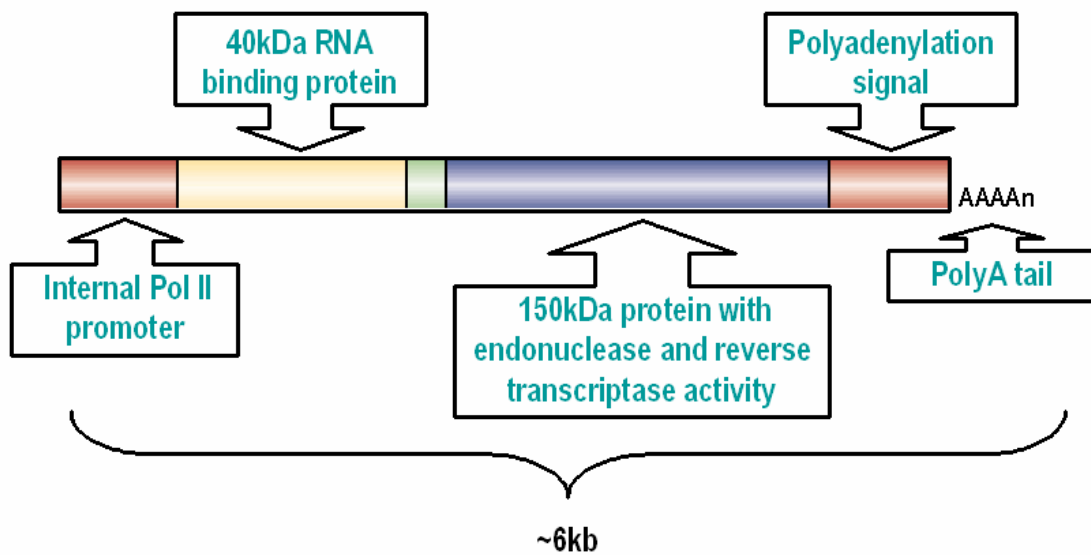
# 1. INTRODUCTION

## 1.1. Biology of Retrotransposons

Less than 5% of the sequenced human genome is composed of coding sequences while various types of repetitive sequences make up an astounding 50% of the genomic mass (Lander et al. 2001). They can be divided into two major groups i.e., tandem repeats (such as variable number tandem repeats, microsatellites and telomeres) and interspersed repeats called transposable elements. Transposable elements are the largest group of repetitive sequences, accounting for 45% of the genome (Lander et al. 2001). Also known as transposons, these sequences maintain their presence in the genome by moving from one location of the genome to another using either a “cut and paste” or “copy and paste” mechanism. DNA transposons move by excising and reintegrating itself into a different location of the genome. A more successful class of transposons is the retrotransposons, which utilizes a replicative mechanism of amplification via an RNA intermediate. LTR retrotransposons, Long Interspersed Elements (LINEs) and Short Interspersed Elements (SINEs) belong to this group.

LINEs are found abundantly in mammalian genomes, comprising 21% of the human genome (Lander et al. 2001). Their origin can be traced back to the earliest emerged eukaryotes (Malik, Burke, and Eickbush 1999). They are autonomous elements, meaning they can encode their own enzyme machinery required for their transposition. A fully functional LINE is about 6kb in length and contains a 5' untranslated region (UTR) with an internal RNA polymerase II (PolII) promoter activity, two nonoverlapping open reading frames which are separated by an intergenic spacer, a 3' UTR with a

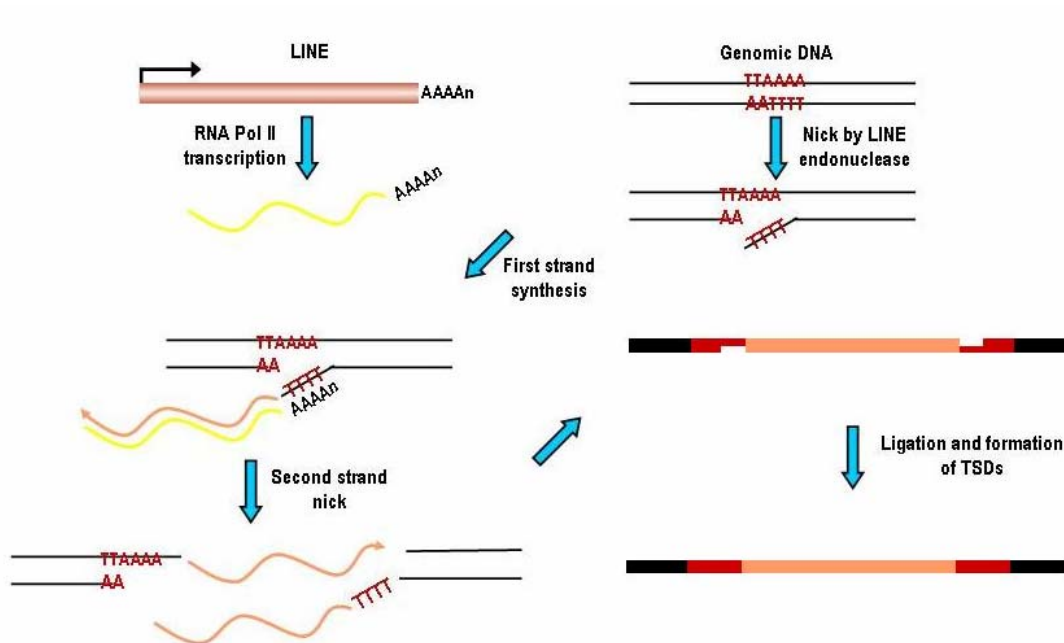
polyadenylation signal, and ends with a polyA tail (Kazazian and Moran 1998) (Figure 1.1). LINEs are usually flanked by target site duplications (TSDs) of about 7 to 20 bp at each end as a result of the retrotransposition process (Fanning and Singer 1987). ORF1 encodes for a 40kDa RNA binding protein (ORF1p) (Martin and Bushman 2001). ORF2 encodes for a 150kDa protein (ORF2p) with an endonuclease and reverse transcriptase activity (Sakaki et al. 1986; Mathias et al. 1991; Luan et al. 1993; Feng et al. 1996)



**Figure 1.1.** Structure of a LINE.

LINEs transpose using a process called target primed reverse transcription (TPRT) (Luan et al. 1993) (Figure 1.2). First, transcription of the element is initiated by the internal promoter. Although LINE transcripts have the same properties as PolIII transcripts, there has been evidence that some of the PolIII machinery may be involved in LINE transcription (reviewed in Deininger and Batzer 2002), suggesting that LINEs are transcribed by a combination of PolIII and PolIII. Translation of the bicistronic LINE mRNA produces ORF1p and ORF2p. The role of ORF1p in retrotransposition is unclear.

However, when the protein is mutated, the efficiency of retrotransposition is greatly reduced, thus showing that ORF1p is essential for LINE mobility (Esnault, Maestre, and Heidmann 2000; Wei et al. 2001). The endonuclease activity of ORF2p makes a single strand nick in the host DNA, usually at a 5'-TTTT/AA-3' consensus, generating a short stretch of oligo-Ts with a free 3'OH. This oligonucleotide primes with the polyA tail of the LINE mRNA, which serves as a template for reverse transcription. A second nick is generated in the opposite DNA strand, followed by the synthesis of the complimentary LINE mRNA strand and integration of the LINE cDNA into the genome. The mechanism for this step has yet to be elucidated. Finally, TSDs are generated at each end of the LINE, marking the boundaries of its insertion.



**Figure 1.2.** LINES replicate by a process called Target Primed Reverse Transcription

Three different families of LINEs exist in the human genome i.e., LINE1 (L1), LINE2 (L2) and LINE3 (L3) (Lander et al. 2001). L1 is the youngest and the only actively retrotransposing family. L1 LINEs emerged around 120 mya and are found in all mammals (Smit et al. 1995). There are over 500,000 copies of L1s in the human genome, making up 17% of the sequences. However, a vast majority of the L1s has lost their ability to mobilize due to 5' truncations, 5' inversions and accumulation of deleterious point mutations. Most L1s are truncated at their 5' end with an average size of 900bp, due to the inefficiency of the reverse transcription process. Five prime inversions occur as the result of twin priming, disrupting the arrangement of the ORFs (Ostertag and Kazazian 2001). Only 3000-4000 of all L1s are full length (Kazazian and Moran 1998) and of those, an estimate of only 80-100 L1s is still capable of retrotransposition (Brouha et al. 2003). The observed pattern of retrotransposition is consistent with the “master gene” model of retrotransposition (Deininger et al. 1992). According to this model, only a handful of L1s are responsible for the majority of the LINE retrotransposition in the genome. Mutations that are accumulated on these “master genes” are passed down to their progenies, creating different subfamilies based on the different diagnostic mutations that they share. By tracing the history of mutation events and determining the distribution of the individual L1s in the genome, the age of the subfamilies can be determined. L1Hs preTa and Ta are the youngest subfamilies of L1 in the human genome (Myers et al. 2002; Salem et al. 2003). The L1Hs preTa subfamily has an ACG diagnostic mutation at the 5930–5932 position while the L1Hs Ta subfamily has an ACA diagnostic mutation at the same position. The average ages for these subfamilies are 2.34 and 1.99 million years,

respectively. Since the divergence between humans and chimpanzees is about 4 million years, L1Hs is only found in human.

The LI enzymes exhibit a strong *cis* preference, meaning they usually bind to the mRNA that produces them (Wei et al. 2001). This prevents competition from other cellular RNA and nonfunctional L1 RNA which have polyA tails (Kazazian and Moran 1998). Pseudogene formation has been shown to occur at a much lower rate than the L1 retrotransposition by *trans* interaction with the L1 proteins (Esnault, Maestre, and Heidmann 2000; Wei et al. 2001). Alternatively, L1s could be more deleterious to the genome and undergo negative selection. Boissinot, Entezam, and Furano (2001) reported that full length L1s are more abundantly found on the Y chromosome compared to other autosomes, demonstrating that deleterious L1s are removed by recombination.

## 1.2. The Effects of LINEs on the Human Genome

LINE insertions have contributed to a significant portion of the human genomic mass. Retrotransposons represent about 1/3 of the human genome compared to <1% of the *Drosophila* genome (Kazazian 2004), creating a large genetic load on our genome. Several documented cases of diseases are caused by gene disruption by LINE insertions (Kazazian 2004). They can also create genomic deletions upon their insertion (Gilbert, Lutz-Prigge, and Moran 2002). LINEs are able to limit their negative effects by making truncated nonfunctional RNA copies of themselves using premature polyadenylation signals (Perpelitsa-Belancio and Deininger 2003). Full length L1s could also be removed by recombination to limit their deleterious effects. Ironically, L1 insertions have been proposed to play a role in repairing double stranded breaks by an endonuclease-independent mechanism of transposition (Morrish et al. 2002).

The presence of LINEs can influence the rearrangement of the genome through exon shuffling (Moran, DeBerardinis, and Kazazian 1999). L1s possess a weak polyA signal, which may be missed by PolIII, causing transcription to end at a later polyA signal. Pickeral et al. (2000) estimated that 15% of full length L1s undergo 3' transductions, contributing to 1% of the human genome. Three prime transductions can play an important role in the evolution of new genes. It allows the duplication of exons and genes and aids the formation of new fusion genes upon integration in a new locus. New genes with important novel functions could be assembled via this mechanism.

LINEs also facilitate unequal recombination by providing hotspots for homologous recombination since they are very similar in their sequences. Unequal L1/L1 crossing over can produce deletions or duplications in the genome. Diseases caused by L1/L1 recombination have been reported, although they occur much rarer than *Alu/Alu* recombination-mediated diseases (Deininger and Batzer 1999).

### 1.3. Retrotransposons as Phylogenetic Tools

The use of LINE (Long INterspersed Elements) and SINE (Short INterspersed Elements) insertions as phylogenetic and population genetic markers is increasing. Some recent applications of mobile elements to phylogeny include the elucidation of hominid phylogeny (Salem et al. 2003), an examination of interfamilial relationships in turtles (Sasaki et al. 2004), and clarification of cichlid phylogenetics (Terai et al. 2004). Mobile elements make excellent phylogenetic and population genetic markers primarily because they have two major advantages over more traditional molecular data such as mitochondrial and nuclear sequences sequence.

First, the presence of a mobile element in an individual is thought to represent identity by descent (IBD), since the probability that two different mobile elements would integrate independently in the same chromosomal location is small (Batzer and Deininger 1991; Batzer et al. 1994; Okada et al. 1997; Batzer and Deininger 2002). Once a SINE or LINE inserts into a chromosomal locus it may generate new copies, but there is no evidence that it is ever completely excised or lost from that locus. Polymorphic mobile element insertions should thus reflect population and species relationships more accurately than many other genetic markers (i.e. sequence data, restriction fragment length polymorphisms (RFLP), and microsatellites) in which the sharing of the same allele by two individuals may reflect identity by state only. Each new integration is therefore a record of a unique transposition event that occurred only once in the evolution of a group. A second advantage of these genetic markers is that the ancestral state of an insertion polymorphism is known to be the absence of the element at a particular genomic location (Batzer and Deininger 1991; Perna et al. 1992; Batzer et al. 1994). Precise knowledge of the ancestral state of a genomic polymorphism allows us to draw trees of population and phylogenetic relationships without making unnecessary assumptions (Perna et al. 1992; Batzer et al. 1994; Batzer et al. 1996; Stoneking et al. 1997; Watkins et al. 2001; Batzer and Deininger 2002; Watkins et al. 2003).

This does not mean that mobile elements are without problems with regard to phylogenetic analysis. It is known that insertion homoplasy can occur across distantly related taxa as a function of evolutionary time and variable retroposition rates among species (Hillis 1999; Miyamoto 1999; Cantrell et al. 2001). This can limit the application of mobile elements in examinations of more diverse taxa. Random sorting of

the ancestral allelic lineages, sequence convergence, and sequence exchanges between alleles or duplicated loci have also been identified as likely factors confounding the interpretation of the interrelationships among species. These sorts of events should be relatively rare, however, and sufficient sampling would easily overcome the problems they might introduce.

#### 1.4. Goal of Study

The purpose of this study is to investigate the claims that mobile elements are vulnerable to homoplasy. It is important to determine the frequency of homoplasy events because they can affect the accuracy of phylogeny inference. We conducted PCR to amplify human-specific LINEs in human DNA samples and orthologous sites in a variety of non-human primates. These LINEs are expected to be absent in all non-human primates, so smaller sized PCR products should be detected in the orthologous sites of the non-human primates. Any PCR products that differed from their expected sizes were analyzed by sequencing to determine the nature of the variation.

Recently, Salem et al. (2003) surveyed primate genomic variation at LINE 1 preTa loci. In the course of their study, they noted several instances of potential homoplasy as evidenced by PCR analyses. Such examples have also been noted in another recent examination of the Ta subfamily (Vincent et al. 2003). In that work, no instances of PCR amplification patterns with the potential to be interpreted as homoplasy were due to secondary LINE insertions. Instead, most anomalous amplifications were due instead to parallel insertions of *Alu* elements in the same 100 bp pre-integration site. This present study expands on the work of Salem et al (2003) by examining L1 preTa orthologous loci in a larger non-human primate phylogenetic panel.

## 2. MATERIALS AND METHODS

### 2.1. DNA Samples

Primate DNA samples were isolated from cell lines from Coriell Cell Repositories: HeLa (ATCC# CCL-2), common chimpanzee (*Pan troglodytes*) Wes (Repository# AG06939), pygmy chimpanzee (*Pan paniscus*) (Repository# AG05253), gorilla (*Gorilla gorilla*) Lowland Gorilla (Repository# AG05251), orangutan (*Pongo pygmaeus*) (Repository# NG12256), green monkey (*Cercopithecus aethiops*) (ATCC# CCL-70), owl monkey (*Aotus trivirgatus*) (ATCC# CRL-1556) and galago (adenovirus 12 SV40-transformed *Galago senegalensis* fibroblasts). Sources for the cell lines are listed in Table 2.1.

**Table 2.1.** List of DNA sources for all species studied.

Species	Common Name	Origin	ID number
<i>Homo sapiens</i>	Human	ATCC <sup>a</sup>	CCL2
<i>Pan troglodytes</i>	Common chimpanzee	Coriell <sup>b</sup>	AG06939
<i>Pan paniscus</i>	Bonobo	Coriell <sup>b</sup>	AG05253
<i>Gorilla gorilla</i>	Lowland Gorilla	Coriell <sup>b</sup>	AG05251
<i>Pongo pygmaeus</i>	Orangutan	Coriell <sup>b</sup>	NG12256
<i>Cercopithecus aethiops</i>	Green monkey	ATCC	CCL-70
<i>Aotus trivirgatus</i>	Three-striped owl monkey	ATCC	CRL1556
<i>Galago senegalensis</i>	Senegal galago	Cell line	<sup>c</sup>

<sup>a</sup> From cell lines provided by the American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108.

<sup>b</sup> Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103.

<sup>c</sup> Adenovirus 12 SV-40-transformed fibroblast cell line maintained in the laboratory of Dr. Mark Batzer, Louisiana State University.

<sup>d</sup> Frozen Zoo, San Diego Zoo Research, PO Box 120551, San Diego, CA 92112.

### 2.2. PCR and Gel Electrophoresis

Two hundred fifty-four individual preTa L1 insertion loci were amplified to test for the presence or absence of the elements in individual primate taxa by PCR. The panel

tested included all species listed in Table 2.1. Twenty-five µl PCR amplifications were performed under the following conditions: 10-100ng of template DNA, 40 pM of each oligonucleotide primer 50 mM KCl, 10 mM Tris–HCl (pH 8.4), 1.5 mM MgCl<sub>2</sub>, 0.2mM deoxyribonucleotides, and Taq DNA polymerase (1U). All reactions were subjected to an initial denaturation step at 94°C for 150s, followed by 32 cycles of PCR, each cycle has a 150s denaturation step at 94°C, a 15s primer annealing step at the specified annealing temperature, and a 60s extension step at 72°C, followed by a final 180s extension step at 72°C. Gel electrophoresis was performed on the PCR products on 2% agarose gels stained with ethidium bromide. PCR bands were detected using UV fluorescence.

Primer sequences and annealing temperatures used for PCR amplification have been previously reported Salem et al. (2003). Lack of consistent amplification in non-human primate taxa prompted the redesign of some primers as noted in Table 2.2.

**Table 2.2.** Redesigned preTa L1 oligonucleotide primers.

Name	Forward primer	Reverse primer	PCR product sizes			
			Filled	Empty	Subfamily	AT
LIAD1	TTCCCTTCCTTTGGAATGTCT	ATATGGCCATCTTGACTCAGTG	6778	191	574	54
LIAD5	TCATCTCACAGAGCTCACAG	CTAGGAATCCTTCTGTCTGG	749	326	150	55
LIAD8	TGGTTTCAATCCCTACTTCTGG	TCTGGGTGGAATGATAAAGTCA	985	No TSDs	353	54
LIAD44	ACATTGGTGTCTGAGTGTCTGG	GTTGCTCCAAAGGAACCTTGT	6394	334	314	55
LIAD59	GCAGTGGGACATTGACTCCTAC	TGTGGCATAGGTTTCTGGAAGT	767	No TSDs	286	55
LIAD63	GCCCACTAGTCTGTTTTGTGCT	ATGCCTTGGACATGGTGAATAG	6389	329	254	54
LIAD64	TCTGGAGGCCACTGCTAATC	AGAAAGGCATGACAGCCAGT	6290	236	179	54
LIAD65	TCCTCAAAGTTGATGCTCCTC	CCTTCCCTTGTCCCTCATT	6808	292	573	60
LIAD68	GCATGCATACTGGACAAAACAT	CTTACTTCATCCCATGCACCAC	788	605	201	54
LIAD70	TGTGTTTCAGTATGCGGGTTC	ACAATTTGTGGGCTTAGCAC	959	No TSDs	309	NA
LIAD74	TTTGTTTCAAGCCAATGCTG	TGGCATACTCGTATTCTAAGTGC	1237	No TSDs	310	54
LIAD96	GGGAAAGCTCATTGTTTTCG	ACATCCAGACCACCAGGAAG	1003	169	303	54
LIAD123	TCATTACAGACCATTGACATGC	TAAATGCATTGGCACCACAC	555	243	282	54
LIAD138	CCAGACAAGTTTGCCTTATGACT	TGCTCCCTTATGACCACTATCAT	1109	404	408	54
LIAD149	AGGCCGAAACACAGATAAGC	TGGTGGTGCCCATATTTGTA	2193	541	277	55
LIAD174	CCATTGGACTTCCCTCTTGA	AGAGCTGCACCCAGAGAT	341	159	127	54
LIAD177	CTGCCTTCTGATTGCTCGT	CCTGGGAATTCATCGTAAC	6713	599	684	53
LIAD185	CTAGGGCCACTGGTAGGGTA	ATCACCGGGCTGTAAACAAC	6402	No TSDs	326	55
LIAD186	GTCAGCAGGCCAATGTGAC	GGAGAACTTGCGCATTAGGA	791	138	247	54

(Table cont.)

LIAD190	TTGGGGAATAATCATGCACTC	AAGCAGCATCTACAGGCAAAG	440	228	337	54
LIAD192	AGCAACAGTAAGTCCCCATT	TGACTTTAGTGACTCCTGCTCTTT	1034	227	363	54
LIAD193	TCTTTTACTCCAAAAGGAA	TTGGGTAGATGAAGATGACC	1833	193	236	60
LIAD195	TGTCCACCAGTCCTGTGATG	TGCCTCTGGTAACCGTAGC	6449	341	402	NA
LIAD197	GATTACGGGAAGTTAGGTAGCC	ACCCAGGTACACACACTAGT	6262	185	300	54
LIAD209	AATGAGTTCACGCATTGTGTT	AGCAAAGCAAGGCAGGTATG	1653	200	248	54
LIAD214	AAGTGACGCACCTTCTGCTT	TGGAGGTGACTCCGATGTA	6325	193	344	55
LIAD219	GCAGGGAATATTTGGGACAT	CAGTCCCCACCACACTAGAA	6424	360	395	54
LIAD226	GCCCCTAGAGGCATTTGAGT	CAACAAGTTTACGCAGAACACTG	733	No TSDs	283	54
LIAD242	TGAGAGGGGGATTATTTTGA	CCTAACAGTCAGGAAAGCTGA	6245	186	197	54
LIAD243	GTGAACATGACTGTGATATTTAAGG	TTGTGGTGTGACTGCATGA	1755	131	282	54
LIAD244	CCCCTGTGGTCTTCCTCA	CCAGAGTCTGATGCGTTTGA	591	No TSDs	243	54
LIAD254	CCAAACTTTAAGAACGCCATGT	GTGGGGGAGGTTTAGGGTAG	904	No TSDs	355	54
LIAD261	CTATGGACCACATCTGACTGT	AGTTATTAACCGGCCACTA	6269	222	245	55
LIAD276	CCTAGCCATATTGAACCGTGA	TGGATTTC AAGAGGAGACCAA	727	No TSDs	293	54
LIAD287	TGCCTAAGCCAAATCTGAA	TTCAAATCTCCTCACCTATGG	369	128	177	55
LIAD290	CTCCCATGCCTCAACATCTC	GAACCCACGAGGTTGTAGC	779	204	240	54
LIAD291	TGGAAAAATATCCCATAAATGA	TTTCAGATGGTTTTCAACA	6277	180	311	53
LIAD293	AGCACTTTCTTTGCCTTGGA	TTCACATTCCAGTAGGGGAGA	2184	182	292	54
LIAD295	CCATTCCGCATGGAAAATTA	GCAGCTTTGTACCGAAGTCC	991	No TSDs	327	53
LIAD299	AGAGTTTCCAGCTGCACTC	AGATCAATGGCTCTGCGAAC	6344	No TSDs	280	54
LIAD301	AAATTCCTGAGCGCTAACCT	TGGAATGTGAGGATGAAGGA	1279	157	230	54
LIAD318	ACCTTGACCATGGGATGAAC	ATGCCTGTGGACTTGCTACA	634	179	258	54
LIAD325	AGAATGGATGTTGGGTGCTC	TGTCCCCATGAACAAATTC	1099	No TSDs	358	53
LIAD327	AAAACATATTTGGAGGAGCA	GTGACCTGGTGTITTTGTCT	6315	202	314	55
LIAD328	TGGCCATTCTCATGTTCTCA	AGCATCACCAACACAACCTG	2431	233	240	54
LIAD334	TTGACTTGTTTAGAAAGGGATT	GGATAAAGCTGAAAGCTCAA	6322	233	215	53
LIAD338	TCCAATTTGCAACAGCTACA	CTGCACATTGCTTTGGACAT	6454	186	440	54
LIAD339	GTTAAAAATGCCAGGCTGAT	TGAGAAATGTGTTCTCCAAA	1169	136	349	53
LIAD348	AGGAAGATTCACAGAATGTGA	AGAGTTTGACAAGTGGCTGGT	1379	No TSDs	224	54
LIAD349	ACAAGCTGCAATTGTGTTGG	ACTGCCCTGCTCTCCTTTCA	1805	177	235	54
LIAD355	CTGAGTGCCTGCAATCCTTT	GAAACTGGGTAAACCCCAAG	1692	217	308	54
LIAD359	AAGGGCATATAAAAAGTGGT	GCACCCATTAATCATCATT	6460	356	328	NA
LIAD372	TCGAAATACACTTACGCCTCAA	GGATAAACCAATAAGTGACCATC	2217	172	279	55
LIAD373	GGAGAGGCAAGAAACTCCAA	CTGCACTGTGTGTCATTGGT	850	192	236	55
LIAD383	TGGTGGTCTCAGAGTAAACA	ACCCAAAACATCATTAGTGC	1642	117	1026	54
LIADY8	TCACACGTATCCCTTTGCAG	GCGCTTTGTGTCCTATGTTG	2041	343	432	NA

### 2.3. Cloning and Sequencing

PCR products purified directly or from the agarose gel were cloned with the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA) according to the manufacturer's

directions and clones were sequenced using chain termination DNA sequencing on an ABI 3100 automated DNA sequencer (Sanger, Nicklen, and Coulson 1977). DNA sequences were aligned using MegAlign v5.00. All the sequences generated for the project have been submitted to Genbank under accession numbers AY705214 through AY705231.

### 3. RESULTS

Of a total of 254 loci analyzed by PCR, 235 loci amplified empty sites in one or more non-human primates. The positions and sequences of the empty sites were determined by computationally removing the L1 element along with one of the target site duplications (TSDs) (Vincent et al. 2003). PCR was conducted on the remaining 19 loci using the internal subfamily specific primer (5'-CCTAATGCTAGATGACACG-3') and the 3' flanking primer to investigate the possibility that the L1Hs preTa subfamily is much older than suspected and therefore would have the L1 insert in the non-human primate orthologous sites. Twelve of those loci confirmed the presence of an L1 insertion in human but no PCR product was observed at any of the non-human primate genomes. The other seven loci showed no PCR amplification in either human or non-human primates. The absence of an empty site in all non-human primate samples at the 12 loci may suggest that some genomic sequence was deleted upon the insertion of the L1 element (Gilbert, Lutz-Prigge, and Moran 2002; Symer et al. 2002). If this were the case, we would underestimate the size of the empty site in the non-human primates. The absence of PCR products could also be due to technical reasons caused by mutations in the primer binding sequence, preventing PCR amplification.

The number of individual empty site loci successfully amplified across all non-human primates is listed in Table 3.1. The pattern of successful amplifications in the other primates is consistent with their estimated divergence times when compared to the human lineage. Out of a total of 1134 loci amplified, nine of the empty sites did not

match their predicted sizes. Detailed sequence analysis of the PCR products revealed the precise nature of the events that contributed to those anomalies.

**Table 3.1.** Total orthologous loci amplified in non-human primates.

DNA sample	Common Chimp	Pygmy Chimp	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago	
Total loci analyzed	222	224	224	199	132	91	42	<b>1134</b>

There were four instances (L1AD3, L1AD54, L1AD138 and L1AD207) of sequence insertions in the orthologous sites of the non-human primates (Table 3.2). At locus L1AD3, a simple sequence repeat expansion increased the size of the amplicon in green monkey by 111 bp. Segments of the L1AD54 and L1AD138 loci were duplicated in the green monkey and gorilla sequence, respectively. Three deletion events were observed in various orthologous empty sites (L1AD9, L1AD44 and L1AD207). Figure 3.1 illustrates the different types of anomalous empty sites at the L1AD207 locus. These events could either be the result of strand slippage during replication or unequal homologous recombination. The frequency of these events is a good indication that the genome is not static.

**Table 3.2.** Simple sequence insertions and genomic deletions.

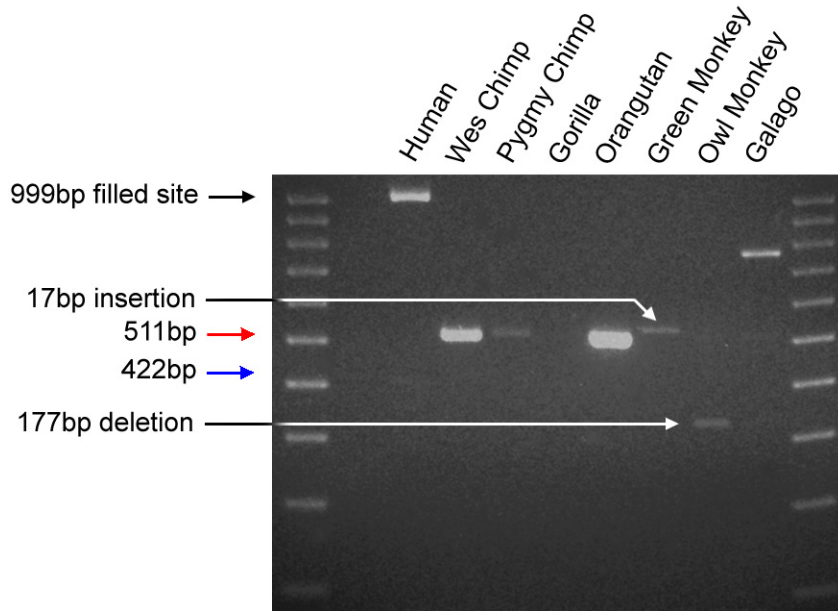
L1Hs preTa locus	Human	Common chimp	Pygmy Chimp	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago
L1AD3	0	0	-	-	-	+111 bp <sup>a</sup>	-	0
L1AD9	-	-	-	-	-	-135 bp <sup>b</sup>	0	0
L1AD44	0	-	-	-	0	-39 bp <sup>b</sup>	0	0
L1AD54	-	-	-	-	-	+40 bp <sup>a</sup>	-	-
L1AD138	0	-	-	+76 bp <sup>a</sup>	-	-	-	0
L1AD207	-	-	-	-	+17 bp <sup>a</sup>	-	-177 bp <sup>b</sup>	-

<sup>a</sup>Simple sequence insertions at orthologous loci

<sup>b</sup>Genomic deletions at orthologous loci

“0” denotes no amplification

“-” denotes presence of the expected PCR product



**Figure 3.1.** Anomalous events in the L1AD207 locus. A 17bp insertion occurred in the green monkey locus and a 177bp deletion in the owl monkey locus independently of the L1AD207 element insertion. An L1-mediated deletion of 89bp of a LIME2 element occurred in the human genomic sequence upon its insertion. The blue arrow denotes the size of the expected empty site (422 bp) determined by computationally removing the L1 insertion. The amplified empty site (511 bp) is denoted by the red arrow.

It is important to note that the deletions mentioned above do not appear to have been facilitated by LINE insertion events. The LINE-mediated deletions occur during the Target Primed Reverse Transcription (TPRT) process and would result in larger empty sites in all orthologous primate genome. The mechanism that leads to such deletions remains unclear. Four such events (L1AD54, L1AD113, L1AD207 and L1AD361) were recovered from our analysis ranging from 14 bp to 89 bp deletions (Table 3.3). All the insertions were found in noncoding region of the genome, mostly deleting portions of other repeated sequences. It is very likely that more LINE-mediated deletions have occurred but have gone undetected by gel electrophoresis because they were too small (1-

20 bp). In addition, some deletions may have been too large (>1kb) to be amplified in a standard PCR reactions.

**Table 3.3.** L1-mediated deletions of the human genomic sequence. Analysis using BLAT and Repeatmasker revealed the nature of the genomic deletions.

L1Hs preTa locus	Size of predicted empty site	Size of observed empty site	Deletion size	Genomic sequence deleted
L1AD54	128bp	149bp	21bp	MERC51C element
L1AD113	80bp	94bp	14bp	Intron of a predicted gene
L1AD207	422bp	511bp	89bp	L1ME2, AluY, MLT1A1 elements
L1AD361	149bp	221bp	72bp	LIMC3 element

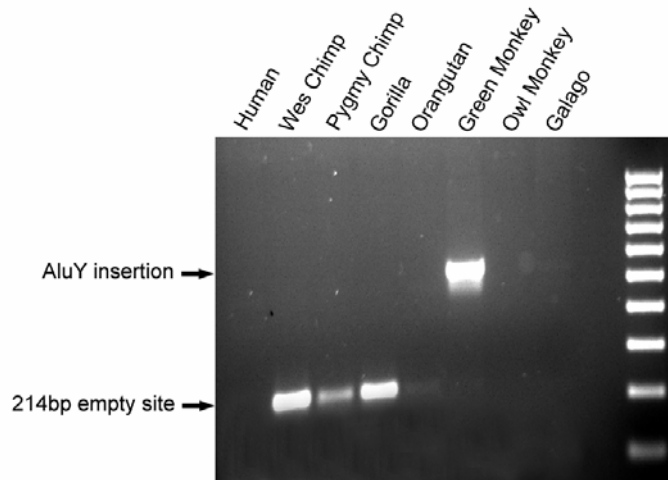
Only two events of independent parallel insertion were detected in all 1134 amplified sites (Table 3.4). The L1AD216 orthologous locus in owl monkey has a truncated L1PA8 insertion just one base away from the L1AD216 insertion site. The proximity of the insertion sites is most likely due to chance alone and does not necessarily indicate the locus is a hotspot for mobile element insertion. An *AluY* element inserted into the green monkey genome 78 bp from the L1AD273 insertion site (Figure 3.2).

**Table 3.4.** Parallel mobile element insertions in non-human primates

L1Hs preTa locus	Human	Common chimp	Pygmy Chimp	Gorilla	Orangutan	Green Monkey	Owl Monkey	Galago
L1AD216	0	-	-	-	-	-	+L1PA8	0
L1AD273	0	-	-	-	0	+ <i>AluY</i>	0	0

However, these parallel insertions do not qualify as authentic homoplastic events for several reasons. Both amplicons were of different sizes compared to the filled sites of the loci in the human lineage. In addition, PCR analysis with the preTa subfamily specific primers did not generate any amplicons, demonstrating that they were not of the L1Hs preTa subfamily. In addition, DNA sequence analysis of the loci revealed that they

did not have the same insertion sites as their human counterparts or identical TSDs upon their insertion. Thus we conclude that they are not authentic homoplastic mobile element insertions.



**Figure 3.2.** Parallel insertion of an *AluY* in the L1AD273 green monkey ortholog. This figure shows a 313 bp *AluY* inserted 78 bp away from the L1AD273 element insertion site. The predicted empty sites were amplified in all orthologous sites except orangutan, owl monkey and galago. No band is visible in human due to the large size of the filled site (>6kb).

#### 4. DISCUSSION

The expansion of mobile elements in the mammalian genome has provided a method to trace the evolutionary history of related taxa. After initial insertion into the genome, a mobile element typically drifts toward being fixed for presence or is more often lost from the population. Fixed LINEs and SINEs remain in the genome and are passed down to all descendants. Therefore, mobile elements that are shared by a group of organisms indicate that they have a common ancestor. This makes LINEs and SINEs ideal markers to examine the evolutionary history of closely related organisms (Nikaido, Rooney, and Okada 1999; Shedlock and Okada 2000; Kawai et al. 2002; Salem et al. 2003). However, the reliability of mobile as markers has been questioned because they may be susceptible to insertion homoplasy (Hillis 1999).

The present data strongly support the hypothesis that individual mobile element insertions are unique events in the evolutionary history of a genome. None of the anomalous preTa amplification patterns resulted from authentic parallel forward insertions. In addition, there was no evidence of the clean deletion of the preTa LINE observed in the human lineage. The chances of a true parallel insertion to occur are so slim that they can practically be ignored.

A true parallel L1 insertion is defined as a secondary insertion of an L1 of the same subfamily and length in the exact same locus in a different genome. The insertion would also have to create identical TSDs as the primary insertion. To date, no parallel mobile element insertions in non-human primate taxa fit these criteria, thus supporting the idea that mobile elements are homoplasy-free markers (Salem et al. 2003; Vincent et

al. 2003). We use the term “parallel insertion” and “homoplasy” loosely in our analysis to describe any secondary mobile element insertions that were amplified by PCR of the orthologous loci. Several examples of apparent mobile element homoplasy have been reported in the past (Cantrell et al. 2001; Roy-Engel et al. 2002; Salem et al. 2003; Salem et al. 2003), although none have ever been reported for L1 elements. Detailed DNA sequence analysis easily showed that all of the events within the primate lineage involving *Alu* repeats were not truly homoplastic insertions. However, it is often impractical to sequence all or some of the amplified PCR products to determine their authenticity, thus apparent homoplasy can pose a problem when mobile elements are used as phylogenetic markers. Therefore, it is therefore necessary to estimate the frequency of apparent homoplasy.

The frequency of observed anomalous empty sites in the gorilla, orangutan, green monkey and owl monkey are 0.446% (1/224), 0.503% (1/199), 3.788% (5/132) and 2.198% (2/91), respectively. No anomalous empty sites were recovered from the chimpanzee and galago genomes. Humans diverged from chimpanzees relatively recently about 4 mya (Miyamoto, Slightom, and Goodman 1987), giving them little time to accumulate new insertions in their genome (Hedges et al. 2004). Therefore, the lack of homoplasy is not surprising. Owl monkey and galago, having had a longer divergence time, were expected to have more anomalous empty sites since the chances of insertion homoplasy increase across more distantly related taxa. Our observations were probably skewed because many of the owl monkey and galago loci could not be amplified by PCR. Sequence divergence between human and more distant taxa is the most likely cause of

this lack of amplification. Thus, the number of anomalous events may be an underestimate, especially in owl monkey and galago.

From our analysis, we calculated an overall regional parallel insertion frequency of 0.176% (2/1134). The individual parallel insertion rates for green monkey and owl monkey are 0.758% (1/132) and 1.099% (1/91), respectively. Similar frequencies were previously reported by Vincent et al. (2003). Taking their data into account, a total of seven *Alu* and one LINE parallel insertion events were recovered from an analysis of 2470 orthologous loci. This reflects the differential amplification dynamics of *Alu* and L1 elements in the primate lineage.

Gene conversion is another mechanism that can contribute to homoplasy. Mobile elements of the same family preserve much of their sequence similarity, even after millions of years. These elements create hotspots for gene conversion. Instances of mosaic *Alus* and L1s, which share new and old diagnostic mutation sites, have been recovered (Roy et al. 2000; Roy-Engel et al. 2002) and are thought to be products of gene conversion. An L1 can bind to another L1 and replace part of the target sequence with its own. If a younger L1 is gene converted by an older L1, or vice versa, PCR amplification of those loci would generate full length L1 products in all orthologous sites. No instances of gene conversion were noted in our analysis. This finding is in agreement with the analysis done on the L1Hs Ta subfamily by Vincent et al. (2003). The absence of gene conversion is not surprising because it is a rare event. Since L1s typically have fewer copies and are more dispersed in the genome compared to *Alu* elements, they have a lower rate of gene conversion.

It is possible for L1-mediated deletions to cause diseases if parts of a gene are removed. Since most of the genome is non-coding sequences while less than 5% make up genes, the probability of a mobile element inserting within a gene and causing a deletion is extremely low. Also, most mobile element insertions occur in germ cells (Ostertag et al. 2002). Deletion of genes by a mobile element insertion at this stage may be lethal to the development of the fetus and the mutant phenotype will not be seen. 30.77% (4/13) of the anomalies we found were L1-mediated deletions. This is higher than the estimate of 21.62% made by Gilbert, Lutz-Prigge, and Moran (2002). However, since small deviations in PCR product sizes are difficult to detect by our methods, there may be other short L1-mediated deletions (<20 bp different from the predicted empty site size) that have gone undetected. Not surprisingly, most of the L1-mediated deletions occurred in other mobile elements since they are the most predominant in the genome.

L1-mediated deletions can cause homoplasy if the size of the deletion is similar to the size L1 insertion. PCR amplification would result in filled and empty sites of the same size, resembling a parallel insertion in each primate orthologous locus. No such events were noted in our study, confirming that the likelihood of such an event to occur is rare.

The lack of homoplasy observed in LINES can be explained in several ways. Since LINES have a slower amplification rate than *Alu* elements, the probability of a parallel insertion is lower. In addition, new L1 insertions are variable in length due to 5' truncations, which are easily distinguished by PCR as different size L1 insertions. However, only a fraction (~10%) of *Alu* elements are truncated during the TPRT process, making parallel insertions of the same size more common. Alternatively, parallel L1

insertion could be harder to detect by PCR if they are too large to be amplified. Most L1s are larger than 1kb in size while *Alu* elements are only about 300kb.

All of the anomalies found differed from their expected sizes by more than 10 bp. There were probably more examples of anomalies that were not detected because of their small size differences. This does not affect our study of homoplasmy since *Alu* and LINE insertions are usually 300 bp or larger. Therefore, small anomalies can be safely ignored because they cannot contribute to any homoplasmy events. In all 1134 amplified empty sites, none qualified as authentic homoplasmy events.

## REFERENCES

- Batzer, M. A., S. S. Arcot, J. W. Phinney, et al. 1996. Genetic variation of recent *Alu* insertions in human populations. *J Mol Evol.* **42**:22-9.
- Batzer, M. A. and P. L. Deininger. 1991. A human-specific subfamily of *Alu* sequences. *Genomics.* **9**:481-7.
- \_\_\_\_\_. 2002. *Alu* repeats and human genomic diversity. *Nat Rev Genet.* **3**:370-9.
- Batzer, M. A., M. Stoneking, M. Alegria-Hartman, et al. 1994. African origin of human-specific polymorphic *Alu* insertions. *Proc Natl Acad Sci U S A.* **91**:12288-92.
- Boissinot, S., A. Entezam and A. V. Furano. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol Biol Evol.* **18**:926-35.
- Brouha, B., J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V. Moran and H. H. Kazazian, Jr. 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A.* **100**:5280-5.
- Cantrell, M. A., B. J. Filanoski, A. R. Ingermann, K. Olsson, N. DiLuglio, Z. Lister and H. A. Wichman. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. *Genetics.* **158**:769-77.
- Deininger, P. L. and M. A. Batzer. 1999. *Alu* repeats and human disease. *Mol Genet Metab.* **67**:183-93.
- \_\_\_\_\_. 2002. Mammalian retroelements. *Genome Res.* **12**:1455-65.
- Deininger, P. L., M. A. Batzer, C. A. Hutchison, 3rd and M. H. Edgell. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**:307-11.
- Esnault, C., J. Maestre and T. Heidmann. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet.* **24**:363-7.
- Fanning, T. G. and M. F. Singer. 1987. LINE-1: a mammalian transposable element. *Biochim Biophys Acta.* **910**:203-12.
- Feng, Q., J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell.* **87**:905-16.
- Gilbert, N., S. Lutz-Prigge and J. V. Moran. 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell.* **110**:315-25.

- Hedges, D. J., P. A. Callinan, R. Cordaux, J. Xing, E. Barnes and M. A. Batzer. 2004. Differential alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* **14**:1068-75.
- Hillis, D. M. 1999. SINEs of the perfect character. *Proc. Natl. Acad. Sci. U S A.* **96**:9979-81.
- Kawai, K., M. Nikaido, M. Harada, S. Matsumura, L. K. Lin, Y. Wu, M. Hasegawa and N. Okada. 2002. Intra- and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. *J Mol Evol.* **55**:284-301.
- Kazazian, H. H., Jr. 2004. Mobile elements: drivers of genome evolution. *Science.* **303**:1626-32.
- Kazazian, H. H., Jr. and J. V. Moran. 1998. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19**:19-24.
- Lander, E. S., L. M. Linton, B. Birren, et al. 2001. Initial sequencing and analysis of the human genome. *Nature.* **409**:860-921.
- 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.
- 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.
- Martin, S. L. and F. D. Bushman. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol.* **21**:467-75.
- Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke and A. Gabriel. 1991. Reverse transcriptase encoded by a human transposable element. *Science.* **254**:1808-10.
- Miyamoto, M. M. 1999. Molecular systematics: Perfect SINEs of evolutionary history? *Curr. Biol.* **9**:R816-9.
- Miyamoto, M. M., J. L. Slightom and M. Goodman. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. *Science.* **238**:369-73.
- Moran, J. V., R. J. DeBerardinis and H. H. Kazazian, Jr. 1999. Exon shuffling by L1 retrotransposition. *Science.* **283**:1530-4.

- Morrish, T. A., N. Gilbert, J. S. Myers, B. J. Vincent, T. D. Stamato, G. E. Taccioli, M. A. Batzer and J. V. Moran. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet.* **31**:159-65.
- Myers, J. S., B. J. Vincent, H. Udall, et al. 2002. A comprehensive analysis of recently integrated human Ta L1 elements. *Am J Hum Genet.* **71**:312-26.
- Nikaido, M., A. P. Rooney and N. Okada. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. U S A.* **96**:10261-6.
- Okada, N., M. Hamada, I. Ogiwara and K. Ohshima. 1997. SINEs and LINEs share common 3' sequences: a review. *Gene.* **205**:229-43.
- Ostertag, E. M., R. J. DeBerardinis, J. L. Goodier, Y. Zhang, N. Yang, G. L. Gerton and H. H. Kazazian, Jr. 2002. A mouse model of human L1 retrotransposition. *Nat Genet.* **32**:655-60.
- Ostertag, E. M. and H. H. Kazazian, Jr. 2001. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **11**:2059-65.
- Perepelitsa-Belancio, V. and P. Deininger. 2003. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet.* **35**:363-6.
- Perna, N. T., M. A. Batzer, P. L. Deininger and M. Stoneking. 1992. *Alu* insertion polymorphism: a new type of marker for human population studies. *Hum Biol.* **64**:641-8.
- Pickeral, O. K., W. Makalowski, M. S. Boguski and J. D. Boeke. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* **10**:411-5.
- Roy, A. M., M. L. Carroll, S. V. Nguyen, A. H. Salem, M. Oldridge, A. O. Wilkie, M. A. Batzer and P. L. Deininger. 2000. Potential gene conversion and source genes for recently integrated *Alu* elements. *Genome Res.* **10**:1485-95.
- Roy-Engel, A. M., M. L. Carroll, M. El-Sawy, A. H. Salem, R. K. Garber, S. V. Nguyen, P. L. Deininger and M. A. Batzer. 2002. Non-traditional *Alu* evolution and primate genomic diversity. *J Mol Biol.* **316**:1033-40.
- Sakaki, Y., M. Hattori, A. Fujita, K. Yoshioka, S. Kuhara and O. Takenaka. 1986. The LINE-1 family of primates may encode a reverse transcriptase-like protein. *Cold Spring Harb Symp Quant Biol.* **51 Pt 1**:465-9.

- Salem, A. H., G. E. Kilroy, W. S. Watkins, L. B. Jorde and M. A. Batzer. 2003. Recently integrated Alu elements and human genomic diversity. *Mol Biol Evol.* **20**:1349-61.
- Salem, A. H., J. S. Myers, A. C. Otieno, W. S. Watkins, L. B. Jorde and M. A. Batzer. 2003. LINE-1 preTa elements in the human genome. *J Mol Biol.* **326**:1127-46.
- Salem, A. H., D. A. Ray, J. Xing, P. A. Callinan, J. S. Myers, D. J. Hedges, R. K. Garber, D. J. Witherspoon, L. B. Jorde and M. A. Batzer. 2003. *Alu* elements and hominid phylogenetics. *Proc Natl Acad Sci U S A.* **100**:12787-91.
- Sanger, F., S. Nicklen and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U S A.* **74**:5463-7.
- Sasaki, T., K. Takahashi, M. Nikaido, S. Miura, Y. Yasukawa and N. Okada. 2004. First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: an example from the turtle superfamily Testudinoidea. *Mol Biol Evol.* **21**:705-15.
- Shedlock, A. M. and N. Okada. 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays.* **22**:148-60.
- 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.
- Stoneking, M., J. J. Fontius, S. L. Clifford, H. Soodyall, S. S. Arcot, N. Saha, T. Jenkins, M. A. Tahir, P. L. Deininger and M. A. Batzer. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**:1061-71.
- Symer, D. E., C. Connelly, S. T. Szak, E. M. Caputo, G. J. Cost, G. Parmigiani and J. D. Boeke. 2002. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell.* **110**:327-38.
- Terai, Y., N. Takezaki, W. E. Mayer, H. Tichy, N. Takahata, J. Klein and N. Okada. 2004. Phylogenetic relationships among East African haplochromine fish as revealed by short interspersed elements (SINEs). *J Mol Evol.* **58**:64-78.
- Vincent, B. J., J. S. Myers, H. J. Ho, G. E. Kilroy, J. A. Walker, W. S. Watkins, L. B. Jorde and M. A. Batzer. 2003. Following the LINES: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. *Mol Biol Evol.* **20**:1338-48.
- Watkins, W. S., C. E. Ricker, M. J. Bamshad, M. L. Carroll, S. V. Nguyen, M. A. Batzer, H. C. Harpending, A. R. Rogers and L. B. Jorde. 2001. Patterns of ancestral

- human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. *Am J Hum Genet.* **68**:738-52.
- Watkins, W. S., A. R. Rogers, C. T. Ostler, et al. 2003. Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Res.* **13**:1607-18.
- 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**:1429-39.

## **VITA**

Huei Jin Ho was born in Penang, Malaysia, on May 1<sup>st</sup>, 1983, the son of Kah Keong Ho and Gooi Chee Thow. He began his college education in Kolej Damansara Utama in Penang, Malaysia, in Spring 2000 and a year later, he transferred to Louisiana State University in Baton Rouge, Louisiana. He received a Bachelor of Science degree in biological sciences from Louisiana State University in December 2002. He is currently completing his Master of Science degree in biological sciences in Louisiana State University and will graduate in December 2004. His future plans include obtaining a PhD in Cornell University and specializing in cancer research in hope to make a significant contribution to the scientific community.