2018

Evolutionary History of the mysTR Retrotransposons and Connection to mysERV Elements

Tessa Passarelli

Follow this and additional works at: https://commons.emich.edu/honors

Part of the Biology Commons

Recommended Citation
Passarelli, Tessa, "Evolutionary History of the mysTR Retrotransposons and Connection to mysERV Elements" (2018). Senior Honors Theses. 582.
https://commons.emich.edu/honors/582
Evolutionary History of the mysTR Retrotransposons and Connection to mysERV Elements

Abstract
Retrotransposons are a group of transposable elements (TEs) that integrate into new chromosomal locations via an RNA intermediate. They are known as "jumping genes" and are transcribed as RNA, copied into DNA, and then moved to new sites in the genome. These elements comprise a large proportion of mammalian genomes. There are two types of retrotransposons—those with or without long terminal repeats. The most extensively studied family of LTR elements is known as endogenous retroviruses, which form when retroviruses integrate their own DNA within the genome of their hosts. The mysTR family of LTR retrotransposons appears to be limited to genomes of the Cricetidae family of rodents and is potentially related to mysERV6 elements of the Muridae rodent family. The objective of this study is to trace the ancestry of these elements. By utilizing "intra-mysTR PCR", we generated and analyzed sequences of mysTR and mysTR-related elements from various rodents. The generation of a molecular phylogeny using sequences from nutria, guinea pig, squirrel, and deer mouse, generated in this study, along with additional available DNA sequence data suggest the mysERV6 and mysTR elements originated from a single retroviral integration. These results provide continued insights into the origins and evolution of mammalian TEs, particularly since different ERVs are active in certain rodent genomes.

Degree Type
Open Access Senior Honors Thesis

Department
Biology

First Advisor
Dr. David Kass

Second Advisor
Dr. Aaron Liepman

Third Advisor
Dr. Marianne Laporte

Keywords
Rodents, LTR Retrotransposons, Cricetidae, Mammalian Genome

Subject Categories
Biology

This open access senior honors thesis is available at DigitalCommons@EMU: https://commons.emich.edu/honors/582
EVOLUTIONARY HISTORY OF THE MYSTR RETROTRANSPOSONS AND
CONNECTION TO MYSERV ELEMENTS

Tessa Passarelli
A Senior Thesis Submitted to the
Eastern Michigan University
Honors College
in Partial Fulfillment of the Requirements for Graduation with Honors in Biology

Approved at Ypsilanti, Michigan, on this date May 4, 2018

Supervising Instructor (Dr. David Kass)

Honors Advisor (Dr. Aaron Liepmann)

Department Head (Dr. Marianne Laporte)

Honors Director (Dr. Ramona Caponegro)
Acknowledgments

I would like to thank Dr. Kass for his time, patience, and hard work to help me write this thesis and the opportunity to work in his lab. I would also like to thank the Eastern Michigan University Honors College for giving me this amazing opportunity for research and the ability to share my research with others.
Abstract

Retrotransposons are a group of transposable elements (TEs) that integrate into new chromosomal locations via an RNA intermediate. They are known as "jumping genes" and are transcribed as RNA, copied into DNA, and then moved to new sites in the genome. These elements comprise a large proportion of mammalian genomes. There are two types of retrotransposons—those with or without long terminal repeats. The most extensively studied family of LTR elements is known as endogenous retroviruses, which form when retroviruses integrate their own DNA within the germ line of their hosts. The *mysTR* family of LTR-retrotransposons appears to be limited to genomes of the Cricetidae family of rodents and is potentially related to *mysERV6* elements of the Muridae rodent family. The objective of this study is to trace the ancestry of these elements. By utilizing "intra-*mysTR* PCR", we generated and analyzed sequences of *mysTR* and *mysTR*-related elements from various rodents. The generation of a molecular phylogeny using sequences from nutria, guinea pig, squirrel, and deer mouse, generated in this study, along with additional available DNA sequence data suggest the *mysERV6* and *mysTR* elements originated from a single retroviral integration. These results provide continued insights into the origins and evolution of mammalian TEs, particularly since different ERVs are active in certain rodent genomes.
Introduction

Overview of Transposable Elements

Mammalian genomes consist of unique, moderately repetitive, and highly repetitive sequences (Charlesworth et al 1994). The unique component includes roughly 20,000-30,000 protein-encoding genes of mammalian genomes (Waterston et al 2002). Transposable elements (TEs) comprise the majority of the moderately repetitive DNA, and are generally interspersed throughout mammalian genomes, accounting for up to nearly half of the DNA sequence (Waterston et al 2002). TEs continue to shape mammalian genomes by integrating into new genomic locations. These insertional mutations can lead to gene disruption, deleterious mutations, and chromosome rearrangements (Chenais et al 2012), thereby continuing to impact the genetics of organisms and potentially the evolution of species.

Classes of Transposable Elements

TEs are categorized as either Class I or Class II elements based on their mechanisms of mobility (Agren 2014) (Figure 1). Transposons are representative of Class II elements, with mobility being DNA-mediated, not with an RNA intermediate, and generally move into new genomic locations via a cut and paste mechanism, thereby not leading to a net increase in copy number (Benjak et al 2008). Retrotransposons are representative of Class I elements, which are mobilized via an RNA intermediate that is reverse transcribed to complementary DNA (cDNA), the copy being integrated into a new chromosomal location. This is referred to as replicative dispersal and leads to the net increase in copy number of the TE in the genome (Hu and Hughes 2012).
**Classification of Retrotransposons**

Retrotransposons have been classified into short *interspersed* DNA elements (SINEs) and long *interspersed* DNA elements (LINEs) based on their length, with SINEs being around 100-700 base pairs in length and LINEs going up to 6,000 base pairs (Singer 1982). SINEs seemingly originated when retropseudogenes of tRNA or 7SL RNA were integrated into a host’s genome and passed down through the generations (Kramerov and Vassetzky 2011). SINEs are non-autonomous in mobility because they do not encode a functional reverse transcriptase protein sequence or functional enzymes and therefore must rely on proteins encoded by other TEs for retrotransposition (Dewannieux and Heidmann 2005). An extensively studied 7SL-derived SINE is the Alu family from humans, which was formed from a fusion of two similar fossil antique monomers (FAMs) for the two separate arms of the SINE (Quentin 1992). Alu elements continue to be active in the human genome, contributing to a wide array of disorders that include hemophilia, neurofibromatosis, breast cancer, and Apert syndrome (Deininger *et al* 2009).
LINEs comprise roughly 20% of the human genome and are autonomous, encoding the proteins necessary for their own retrotransposition (D’ambrosio et al. 1986). LINEs can be further divided into two categories—those containing long terminal repeats (LTRs) and those without LTRs or non-long terminal repeat (non-LTR) retrotransposons. The LTRs contain up to hundreds of base pairs that flank the LTR retrotransposons in the same orientation (Han 2010). The non-LTR retrotransposons include L1 elements in mammals, of which full-length elements have an RNA polymerase II internal promoter sequence, 5’ and 3’ untranslated regions and two open reading frames (ORFs) that are intact in the few active elements referred to as master or source genes (Martin 2006) (Figure 2). ORF1 encodes an RNA binding protein, whereas ORF2 encodes a protein having reverse transcriptase and endonuclease activity (Martin 2006) (Figure 2). In rodents, these elements cause double stranded DNA break repair and alternative splicing when they are added to the genome (Ostertag et al. 2007). As a result of a weak poly (A) signal, L1 elements can move downstream to a new location and are cleaved after a downstream poly (A) signal, shuffling sequences to new genomic locations, a process referred to as 3’ transduction (Moran et al. 1999) leading to exon shuffling as well as providing regulatory sequences to new genomic sites. These 3’ transduction events cause about 1 in 1,200 human mutations, which makes understanding their mechanisms necessary for understanding their role in diseases (Kazazian 2004). Non-LTR retrotransposons are also active among different orders of mammals, but inactive in a few species such as such as South American rodents (Grahn et al. 2005) and megabats (Cantrell et al. 2008) while various active LTRs retrotransposons are most commonly limited to certain taxonomic groups of mammals (Erickson et al. 2011).
Figure 2. Structures of non-LTR and LTR retrotransposons as shown by the retrovirus Moloney murine leukemia (Mo-MLV). Ty1 of *S. cerevisiae* and gypsy of *D. melanogaster* are LTR retrotransposons, and L1Hs is a non-LTR retrotransposons found in humans. Reverse transcriptase (RT) and nuclease (N) are indicated (Finnegan 1997).

There are various families of LTR retrotransposons- Ty1-copia, Ty3-gypsy, IAP, *mysERV*, *mysTR*, and endogenous retroviruses. Ty1-copia is generally found in single celled algae, angiosperms, or gymnosperms. It encodes four types of proteins: integrase, protease, ribonuclease H, and reverse transcriptase (Wicker and Keller 2007). Ty3-gypsy retrotransposons are also found in the plant kingdom. They can be subdivided into different lineages based on structure, conserved protein sequence motifs, and the presence of specific protein domains (Marín and Llorens, 2000). IAPs are intra-cisternal A-type particles consisting of two open
reading frames flanked by two identical LTRs and have amplified to roughly 1,000 copies in European house mice (Mus) (Ray et al 2011).

The most commonly studied and abundant family of LTR retrotransposons in mammals is the endogenous retroviruses (ERVs) (Havecker et al 2004). ERVs formed when retroviruses integrated their own DNA within the germ line of their hosts (Belshaw et al 2004). HERVs (human endogenous retroviruses) are known as “fossil viruses” that have been transmitted vertically in a Mendelian manner. They contain gag, pol, and env genes between two LTRs (Nelson et al 2003). Though ERVs make up 5-8% of human DNA, they appear to be dormant, as post-integration mutations likely disrupted the coding for functional proteins associated with retrotransposition (Nelson et al 2003; Belshaw et al 2004). In contrast to HERVs, some ERV families in rodents are active (Benit et al 1999). ERV elements are highly active in the rodents of the family Muridae, particularly the intra-cisternal A-type particle (IAP) elements representing one of the more numerous ERVs in murid genomes (Qin et al 2010), offering a unique opportunity to further characterize the origin, evolution, and genetic impact of ERV retrotransposons. The mouse genome is composed of 21% of non-LTR retrotransposons and 8% LTR-retrotransposons (Waterston et al 2002).

*Elements of Cricetid Rodents*

Rodents of the family Cricetidae have an active ERV element, the *mys* element, which has undergone recent extensive amplification in various genera of this family, most notably in the genus *Peromyscus* (Wichman et al 1985). The *mys* element is characterized by features derived from retroviruses, such as long terminal repeats and sequences homologous to reverse transcriptase (Pine et al 1988) (Figure 2). MysERV elements have been previously identified in
rodents found in the family Muridae. With the recent amplification of *mys* elements, it is possible to further understand the role of ERV activity in rodents, most notably in relation to a recently identified ERV found in cricetid rodents known as *mysTR*. There is a hypothesis that certain elements became more active by filling a niche left by "dead" elements (Cantrell *et al* 2005). However, further studies found no reciprocal correlation between *mysTR* activity and L1 activity (Erickson *et al* 2011). Additionally, *mysTR* elements appear to be limited to cricetid rodents (Cantrell *et al* 2005).

*mysTR Elements of Cricetid Rodents*

*MysTR* is a transposable element and another ERV, which may be related to the *mysERV* element found in rodents of the family Muridae (Cantrell *et al* 2005). *mysTR* was first identified in the genome of the L1-inactive *Oryzomys palustris* by a phylogenetic screening technique with the closest match via RepBase to the *mysERV6* element in mice (Muridae) (Cantrell *et al* 2005). *mysERV* elements are common in the genomes of rodents in the Muridae family (Stocking and Kozak 2005). While *mysERV6* and *mysTR* are two different elements, they have some sequence similarities to each other and to beta-like retroviruses (Cantrell *et al* 2005). *mysTR* elements were previously examined using degenerate primers based on conserved regions of betaretroviruses (Cantrell *et al* 2005), suggesting both *mysERV* and *mysTR* may be related by a common ancestor and possibly derived from a common ancestral retroviral integration.

Since retrotransposons continue to shape mammalian genomes and impact the genetics of their hosts, understanding their origins and evolutionary history is expected to provide valuable information regarding their activity over time, as well as their role in the dynamics of mammalian genomes. Evidence of recent activity of retrotransposons includes maintenance of
ORFs, individual elements having high sequence similarity to each other and to a consensus sequence, and identification of recent integration events (Martin 2006). The significance of retrotransposons has been debated, with one theory stating that they help to drive phenotypic variation by their involvement in gene rearrangements via unequal homologous crossovers (Whitelaw and Martin, 2001). Retrotransposons also are thought to affect the evolution of primate genomes due to their high numbers, in turn changing the structure of the genome. However, retrotransposons seem more active in rodent genomes (Knisbacher and Levanon 2015). These elements are also credited with the increasing size of the human genome, with an insertion in about 1/20 births. Besides aiding in creating new arrangements of genes during gestation, the rearrangements of genes are also thought to play a role in cancer (Cordaux and Batzer 2006). Assessing the evolutionary history of $mysTR$ and $mysERV6$ could provide insights into the origin of these seemingly related LTR-retrotransposon families and reveal information on how they relate to each other and where they diverged from one another within mammalian evolution. Additionally, retrotransposons continue to be active and cause insertional mutagenesis. These insertions have been associated with various disorders in humans and rodents (Beck et al 2011), as well as having roles in epigenetics and evolution (Stocking and Kozak, 2005). When retrotransposons are inserted into genomes, they can cause different heritable diseases such as neurofibromatosis, cystic fibrosis, and hemophilia (Hu and Hughes, 2012). Approximately 0.3% of human mutations are due to insertions of retrotransposons into the genome (Cordaux and Batzer 2010). Contrasting TEs in more limited groups of mammals in relation to those that predate the radiation of mammalian species offers the potential to assess features that allow for the relative success of various elements. The objective of this study, therefore, is to understand the evolutionary history of the $mysTR$ family of LTR
retrotransposons, and perhaps explore its relationship to the expansion of other elements including L1 and *mys*. As retrotransposons continue to shape mammalian genomes (Knisbacher and Levanon, 2015), insights into the evolution of the *mys*TR element, as well as in relation to other known elements may advance our understanding in the evolutionary impact of retrotransposons and the relative impact of retrotransposons on each other.

The hypothesis for this study is that *mys*TR and *mys*ERV6 are derived from shared ancestral retroviral integration in an organism predating the evolution of Cricetidae and Muridae rodents, rodent families within the superfamily Muroidea. The alternative hypothesis is they are a result of independent exogenous retroviral integrations in the two rodent lineages. In contrast to previous work that generated molecular phylogenies of isolated elements among rodents, and other mammals using degenerate PCR primers based on internal betatretroviral sequences, this investigation will incorporate the use of primers based on highly conserved sequences of *mys*TR elements determined via an alignment of *mys*TR sequences extracted from the GenBank database that represent seven different genera (Figure 3). We predict that this approach will allow for a more precise determination of the evolutionary history of *mys*TR elements enabling the assessment of whether *mys*ERV6, or potentially other Muridae rodent elements are more related to *mys*TR, and whether *mys*TR elements originated in an ancestor of rodents in the Cricetidae family, or are more limited to certain cricetid subfamilies, or even more limited taxonomic groups.
Figure 3. Alignment of entire *mysTR* sequences found in different clones of gerbil, hamster, and deer mouse (Hofmann, 2015), and from the GenBank database. Those sequences shown in yellow correspond to the primer sequences used in this study.
Materials and Methods

DNA Samples

DNA samples for this study were obtained from various sources. DNA from *Peromyscus* species was obtained from the *Peromyscus* Genetic Stock Center at the University of South Carolina. DNA from other species was previously isolated from tissue samples obtained from the Museum of Southwestern Biology (Table 1).

**Table 1.** Previously isolated DNA samples and stock numbers

<table>
<thead>
<tr>
<th>Stock Number</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK 86784</td>
<td><em>Onychomys leucogaster</em></td>
</tr>
<tr>
<td>NK 43090</td>
<td><em>Reithrodontomys Fulvescens</em></td>
</tr>
<tr>
<td>NK 9652</td>
<td><em>Neotomodon alstoni</em></td>
</tr>
<tr>
<td>MSB 57563</td>
<td><em>Oryzomys palustris</em></td>
</tr>
<tr>
<td>MSB 140165</td>
<td><em>Sigmodon hispidus</em></td>
</tr>
<tr>
<td>MSB 140883</td>
<td><em>Neotoma albigula</em></td>
</tr>
</tbody>
</table>

*Mus musculus* (European house mouse), *Meriones unguiculatus* (gerbil), *Mesocricetus auratus* (golden hamster), *Rattus norvegicus* (Norway rat), *Sciurus carolinensis* (grey squirrel) *Cavia porcellus* (guinea pig) DNA samples were isolated in previous studies (Kass *et al.* 1996, 1997). DNA was isolated from tissue of wild caught *Myocastor coypus* (nutria) and wild caught fox (species not determined) using the Promega SV Genomic DNA Isolation kit. DNA from *Pan troglodytes* (chimpanzee) was provided by Dr. Mark Batzer, Louisiana State University. DNA from *Homo sapiens* (humans), *Octodon degus* (Degu), and *Oryctolagus cuniculus* (rabbit) was isolated in previous studies (Kass *et al.*, 1995, 2009)
Amplification of intra-mysTR PCR products

The polymerase chain reaction (PCR) was used to amplify internal segments of mysTR and mysTR-related elements among various mammalian species. Primers were generated based on highly conserved sequences within the pro-pol region of the mysTR element (Figure 4) from an alignment of elements among various cricetid rodents obtained from the GenBank database and previous clones (Hofmann 2015). DNA of interest (1 µl of 50 ng) was mixed with Fermentas 10x Dream Taq (1x final concentration), 1 µM mysTR F2 primer and 1 µM mysTR R270A primer, 200 µM dNTPs, and Dream Taq DNA Polymerase (1 unit). The PCR cycler was programmed for 2 minutes at 94°C for denaturation, followed by 29 cycles of 30 seconds at 94°C, 30 seconds at 42°C for annealing, 30 seconds at 72°C for extension, and a final extension step of 5 minutes at 72°C. PCR products were separated and analyzed by electrophoresis on a 2% (w/v) agarose gel electrophoresis with a 1X TAE buffer and a Morganville 6x DNA loading buffer added to samples and stained with 0.5x GelRed. The gel was run at 140 V for 1 hour.

Figure 4. Diagram of the mysTR element, arrows indicate location of primers that were used to amplify the highly conserved sequence


**Generation of intra-mysTR PCR DNA Libraries**

Intra-mysTR PCR DNA libraries were generated using the TA Cloning System (Promega). PCR products (3 µl) were directly ligated into 1 µl of pGEMT-Easy plasmid (Figure 5), with 1 µl T4 DNA ligase and 5 µl 2x rapid ligation buffer according to the manufacturer's protocol. Ligation products were then transformed into competent JM109 *E. coli* cells. Competent bacterial cells (50 µl) were mixed with 3 µl of ligation product. The cells were then placed on ice for 30 minutes, heat shocked at 42°C for 30 seconds and placed on ice for 5 minutes. 950 µl of liquid SOC media was added to the transformation and it was placed in a shaking incubator at 37°C for 1 hour at 105 rpm. Transformed bacteria were cultured on LB ampicillin plates coated with 50 µl each of IPTG (100 µl of 100 mM) and X-gal (20 µl of 50 mg/ml) and incubated overnight for blue/white screening. White bacterial colonies were picked. The selected colonies were first patched on a LB/Amp plate as a backup, and then added to a sterile test tube with 1.5 ml LB containing 100 µg/ml of Ampicillin and shaken at 37°C for 12 hours.

**Figure 5.** pGEM-T Easy Vector used in ligation
Isolation of Plasmid DNA

The Promega Wizard Plus SV Minipreps DNA Purification Systems kit was used to isolate the plasmid DNA from the transformed bacteria. The manufacturer’s recommended protocol was used.

Verification of Insert

Restriction enzyme digest

In order for the recombinant DNA to be analyzed to verify the presence and size of an insert in the vector, 17.5 µl of it was cut at EcoRI sites that flank the insert (Figure 5) by digesting with 0.5 µl EcoRI (1 unit) and 1 x CutSmart Buffer at 37°C overnight. The final size should be equivalent to the size of the PCR product, presuming that no EcoRI site exists within the cloned PCR product.

Gel electrophoresis

Agarose gel electrophoresis was used to verify the presence of an insert and was used to determine the size of the insert and if it corresponded to the size of the original PCR product. This was done by adding Morganville 6x DNA loading dye loading to the DNA and loading into an agarose gel stained with 1X GelRed solution, run at 140 V for 1 hour and visualized under UV light using a gel-documentation system. Sizes of the inserts were compared to a 100 bp standard DNA ladder (Morganville Scientific).
DNA Sequence Analysis

The plasmids with expected insert size were submitted for sequencing to the University of Michigan DNA Sequencing Core Facility, and sequenced using a T7 primer. MacVector DNA sequence software was used to analyze the sequences. This included removal of the plasmid and primer sequences and sequence alignments using the ClustalW alignment tool. The sequence of nucleic acids being analyzed was then compared to other *mys*TR sequences isolated in previous studies in this lab (Hofmann 2015) using MacVector to compare the degree of similarity between mammals known to have *mys*TR elements against mammals we suspected to have *mys*TR or *mys*TR-related elements. A molecular phylogeny was generated via ClustalW with the parameters of Open Gap Penalty= 15.0; Extend Gap Penalty= 6.7; and Delay Divergent= 30%.
Results

The determination of the presence of *mys*TR elements was carried out in members of cricetid and non-cricetid rodents. Once the presumptive presence of *mys*TR was determined based on the identification of a PCR product of anticipated size, the PCR products were then cloned and sequenced.

Intra-*mys*TR PCR products from the DNA of deer mouse, chimpanzee, rabbit, nutria, and squirrel were analyzed using gel electrophoresis (Figure 6). A product of expected size was not only observed in the deer mouse (family Cricetidae, suborder Myomorpha) but also in squirrel (rodent suborder Sciuromorpha), nutria (rodent suborder Hystricomorpha), rabbit (order Lagomorpha), and chimpanzee (order Primates). Though bands were notably less intense for chimpanzee and nutria. These results suggest the presence of *mys*TR or *mys*TR-related elements in genomes of these mammals as primers used were strictly based on *mys*TR sequences as opposed to that of betaretroviruses. PCR was repeated with deer mouse as a positive control, along with nutria and guinea pig (also in the suborder Hystricomorpha) (Figure 7). A much more intense band was displayed again for deer mouse, and although not quantitative could indicate fewer *mys*TR elements in Hystricomorpha or amplification of elements simply related to *mys*TR.
Figure 6. Intra-mysTR PCR amplification products among different rodents using pro-pol mysTR primers R270A and F2.
**Figure 7.** Intra-\textit{mysTR} PCR amplification products among different rodents using pro-polv \textit{mysTR} primers R270A and F2.

For confirmation of the presence of \textit{mysTR} or \textit{mysTR} related sequences, the intra-\textit{mysTR} PCR products were directly cloned into the pGEM-T Easy vector to create intra-\textit{mysTR} PCR libraries. DNA from individual clones was isolated and analyzed for inserts (Figures 8 and 9). An insert of expected size was verified for one of three rabbit clones, four of five squirrel clones, and two of three nutria clones (Figure 8) and all three deer mouse clones, both guinea pig clones, and all three nutria clones were also analyzed for inserts (Figure 9).

**Figure 8.** Restriction digest analysis of purified plasmid DNA of potential \textit{mysTR} elements in rabbit and squirrel.
Figure 9. Restriction digest analysis of purified plasmid DNA of potential mysTR elements in deer mouse, guinea pig, and nutria with plasmid and insert indicated.

After confirming the presence of the insert of expected size, the insert sequences were determined. A molecular phylogeny of sequences was generated via Clustal W Alignment (Figure 10). Sequences generated by Hofinann (2015) utilizing an alternative primer set were incorporated in this study. These sequences include the fulvous harvest mouse 1 and 2, marsh rice rat 1, gerbil 1, 2, and 4, and hamster 3 and 4 (Figure 10). The mysERV 6 pro-pol was obtained from Repbase (http://www.girinst.org/repbase/) and trimmed down to the corresponding region (pro-pol) and Pman 10 and 34 were from a previous study (Erickson et al 2011). The mysERV 6 pro-pol sequence clustered with two gerbil sequences, consistent with this family of elements specific for murid rodents. However, one clone isolated from the gerbil (gerbil 4) was highly divergent from the other sequences. Therefore, the sequence was queried in
Repbase using CENSOR (Kohany et al 2006) with the closest match to the rat ERVB6_1_1-RN element indicating this is unrelated to the \textit{mysTR-mysERV6} families of elements. Although the hamster is a cricetid rodent, it is distantly related to the other cricetid rodents used in this study that had previously been determined to contain \textit{mysTR} elements. The squirrel (Sciuromorpha) and nutria (Hystricomorpha) are in different rodent suborders than the other rodents used in the study (suborder Myomorpha) and with minor variations the molecular phylogeny is consistent with known phylogenetic relations of these rodents. Additionally, the one clone isolated from rabbit displays a higher similarity to the \textit{mysTR-mysERV6} family of elements, than does the gerbil 4 clone, suggesting all but gerbil 4 are members of what may be referred to as a \textit{mysTR-mysERV6} retrotransposon superfamily.
Figure 10. Molecular phylogeny of mysTR sequences generated by Clustal W
Discussion

Different families of ERVs have been active in genomes of rodents at various taxonomic levels, such as IAP elements of the Muridae family (Stocking and Kozak 2008; Ray et al 2011) and mysTR and mys of the Cricetidae family. The mysTR elements, which appeared to be limited to cricetid rodents (Cantrell et al 2005), have features and DNA sequence similarities to the inactive mysERV6 element of rodents in the Muridae family (including Old World brown and black rats) (Cantrell et al 2005). Therefore, this study was designed to understand the origin and evolution of these two possibly related elements- mysTR and mysERV. To accomplish this, I isolated potential mysTR and mysTR related elements from cricetid rodents as well as that from a squirrel of the more distantly related Sciuridae family (Sciurognathi suborder), as well as from the guinea pig from the suborder Hystricognathi and rabbit from the order Lagomorpha (Figure 11).

Figure 11. Phylogenetic relationships of rodents used in this study
In contrast to previous studies utilizing degenerative betaretrovial PCR primers, which would indicate the presence of any betaretroviral related element, *mys*TR-specific primers were used in this study to preferentially amplify *mys*TR or *mys*TR-related elements. If *mys*TR is found in other mammals outside cricetids, it shows that it is not limited to cricetid rodents, and could provide clues about the original retroviral integration event. If *mys*TR and *mys*ERV6 shared a common ancestor it would be expected that their evolutionary histories, as depicted by molecular phylogenies of individual sequences, correspond with the phylogenetic relationships of the rodents. If *mys*TR and *mys*ERV arose from independent integration, then this level of correspondence would not be observed. Additionally, if *mys*TR elements showed greater similarity to related elements in more distantly related rodents, such as squirrel, nutria, and guinea pig, then this would be inconsistent with *mys*TR and *mys*ERV6 being derived from a single retroviral integration. Also included was the gerbil, also a murid rodent, as it is possible *mys*TR elements may have become inactive in the rodent lineage leading to mice and rats.

The observance of a correspondence between the molecular phylogeny of *mys*TR, *mys*ERV6, and other *mys*TR-related elements, dating as far back to the rabbit, with known phylogenetic relationships of Rodentia and Lagomorpha supports the hypothesis of a shared origin of *mys*TR and *mys*ERV6. The original source of these families of retrotransposons may have dated back to a common ancestor of rabbits and rodents, and therefore *mys*TR and *mys*ERV6 may be part of a superfamily of elements that diverged over time. This is supported by the one outlier gerbil sequence which corresponds to a different family of rodent ERV elements based on screening the Repbase sequence database.

In conclusion, *mys*ERV and *mys*TR were likely derived from the same ancestral gene, not independent integrations. Over time, sequence variations of the source genes from which copies
of individual elements were generated led to the distinct mysTR and mysERV6 families consistent with the master gene model (Deininger et al 1992) proposed to lead the evolution of distinct subfamilies of Alu (SINE) elements and L1 (LINE) elements, demonstrating concerted evolution. Overall, the data generated in this study have contributed in advancing our knowledge regarding the origin and evolution of retrotransposons. Future studies can incorporate additional sequences from pol-pro regions of other ERVs to verify the mysTR, mysERV6, and mysTR related sequences generated in this study are all part of the same lineage.
Work Cited


