2006

Retrotransposition efficiency of non-autonomous mobile DNA

Erin Radloff

Retrotransposition efficiency of non-autonomous mobile DNA

Abstract

Nearly half of the human genome consists of transposable element (TE) DNA sequences. Most of these sequences were derived from the process of retrotransposition. This term refers to a gene being transcribed into RNA, converted back to DNA and the copy integrated elsewhere in the genome. Short interspersed DNA elements (SINEs) from various rodent species were isolated, using PCR, with the intent of being incorporated into a plasmid construct and analyzed for retrotransposition efficiency. SINE copy numbers vary among rodents; therefore, efficiency can be determined by contrasting copy numbers in respective genomes in different rodents. In future studies, a novel cell culture assay approach will be used to determine the important features and efficiency for retrotransposition.

Degree Type

Open Access Senior Honors Thesis

Department

Biology

First Advisor

David Kass

Second Advisor

James Vandenbosch

Keywords

Mobile genetic elements, Molecular genetics

Subject Categories

Biology

Retrotransposition Efficiency of Non-autonomous Mobile DNA

By

Erin Radloff

A Senior Thesis Submitted to the

Eastern Michigan University

Honors Program

In Partial Fulfillment of the Requirements for Graduation

With Honors in Biology

Approved at Ypsilanti, Michigan on this date: _____

_____

Supervising Instructor: David Kass

_____

Honors Advisor: James Vandenbosch

_____

Department Head: Tammy Greco

_____

Honors Director: Heather Holmes

Table of Contents

Senior Honors Thesis

Abstract

Author: Erin Radloff

Department: Biology

Area: Genetics

Supervising Instructor: David Kass

Honors Advisor: Dr. James Vandenbosch

Title: Retrotransposition Efficiency of Non-autonomous Mobile DNA

Length: 22pp.

Completion Date: 4-21-06

Nearly half of the human genome consists of transposable element (TE) DNA sequences. Most of these sequences were derived from the process of retrotransposition. This term refers to a gene being transcribed into RNA, converted back to DNA and the copy integrated elsewhere in the genome. Short interspersed DNA elements (SINEs) from various rodent species were isolated, using PCR, with the intent of being incorporated into a plasmid construct and analyzed for retrotransposition efficiency. SINE copy numbers vary among rodents; therefore, efficiency can be determined by contrasting copy numbers in respective genomes in different rodents. In future studies, a novel cell culture assay approach will be used to determine the important features and efficiency for retrotransposition.

Introduction

Transposable genetic elements (TEs) are segments of DNA that can be integrated into new genomic locations either through direct DNA transfer (transposons), or, as with rodents, via an RNA intermediate (retrotransposons) (Kass and Batzer, 2003). Class I TEs utilize an RNA mediated process and Class II TEs utilize a DNA mediated process. Both classes include autonomous and nonautonomous elements. Autonomous TEs contain open reading frames (ORFs) that encode the essential transposition proteins. Nonautonomous TEs do not encode proteins, relying on autonomous TEs for the machinery necessary for replication (Kramerov and Vassetzky, 2005).

Class I elements (retroelements) are divided into three groups by their replication strategy and structure: LTR-transposons , Long Interspersed Elements (LINEs), and Short Interspersed Elements (SINEs). LTR-transposons contain long terminal repeats (LTRs) with transcription control sequences and open reading frames (ORFs) whereas LINEs and SINEs both lack LTRs.

LINEs are 4,000-7,000 base pairs (bp) in length. These elements contain an internal promoter sequence preceding two ORFs and a highly polyadenylated region (Kramerov and Vassetzky, 2005). Each of the two ORFs contained within the LINE sequence encode a protein that facilitates the retrotransposition of the mRNA (Dewannieux et al. 2003). The first ORF encodes a single-stranded nucleic acid binding protein. ORF2 encodes a reverse transcriptase with an endonuclease domain (Rinehart et al. 2005). Retrotransposition is a process involving reverse transcription of RNA and subsequent integration into the genome (Vassetzky et al. 2003).

Unlike LINEs, SINEs are non-autonomous. SINEs lack the coding sequence that would otherwise ensure their own retrotransposition. Therefore, SINEs require the enzymes encoded by other elements to ensure transposition (Dewannieux et al. 2003). All SINEs share a number of characteristics, including A box and B box RNA polymerase III promoter sequences, oligo dA-rich 3' regions, and flanking direct repeats indicative of retrotransposition as a primary mode of dispersal (Kass et al. 1994). SINEs can mobilize by the "copy-and-paste" mechanism, in which their own RNA is reverse-transcribed into complementary DNA (cDNA) (Kajikawa et al. 2005). During the creation of a SINE, the 5' part descends from a cellular RNA (usually tRNA); the 3' part is unclear, although the extreme 3' part is similar, in many cases, to the 3' end of conspecific LINEs. Although SINEs are only one type of transposon, the number of their copies makes up one half of the total number of repeated elements in humans and mice (Kramerov et al. 2005). The propagation of SINEs within the genome occurs by the process of retrotransposition.

The predominant family of SINEs in humans is termed Alu elements. There are three major SINE families in the rodent genome referred to as B1, ID, and B2. B1 is derived from a 7SL RNA gene, a component of cytoplasmic ribonucleprotein called *signal recognition particle* involved in translation of secreted proteins in all eukaryotes (Deininger et al. 2003; Vassetzky et al. 2003). B2, ID, and most other SINEs, are derived from tRNA genes (Deininger et al. 2003).

The rat genome contains a 75 nucleotide repetitive element termed ID (identifier sequence) that is present in 1-1.5x10^5 dispersed copies. BC1, an abundant RNA of about 160 nucleotides, which is prevalent in the nervous system, contains sequences

homologous to the ID element (DeChiara et al. 1987). The BC1 gene is transcribed from a conserved, single copy gene in rodents. The gene encoding BC1 RNA represents one of the earliest and possibly the first ID-containing sequence. Based on evidence supporting that BC1 genes showed variations relating to specific changes in its corresponding ID sequence, the BC1 gene is believed to be the master gene responsible for the amplification and evolution of ID elements (Kim et al. 1994).

The amplification of ID elements in the order Rodentia is highly sporadic and does not appear to follow a phylogenetic trend (Kass et al. 1994). The estimated copy numbers of ID elements in the deer mouse, rat and mouse are: *deer mouse genome*: 25,000; *rat genome*: 150,000; *mouse genome*: 4,000; *guinea pig genome*: 200 (Kass et al. 1994). It is this variation in copy number that is of interest when studying the retrotransposition efficiency of the ID.

Retrotransposition is a multiple stepped process resulting in the relocation of a non-LTR element. In the first step, transcription is initiated from an internal promoter localized in the L1 5' UTR. This L1 RNA is then exported to the cytoplasm, where the ORF1 and presumably ORF2 encoded proteins are translated. Both ORF1 and ORF2 encoded proteins (ORF1p and ORF2p respectively) associate with the RNA that encoded them to form a ribonucleic protein particle, which is a proposed retrotransposition intermediate (Gilbert et al. 2005). ORF2p is readily detected in a variety of cultured human and mouse cells. *In vitro* and cultured cell analyses have demonstrated that ORF2p has endonuclease and reverse transcriptase activities that are important for retrotransposition (Gilbert et al. 2005).

The physical action of retrotransposition likely occurs by target-site primed reverse

transcriptase (TPRT). During TPRT, L1 endonuclease cleaves genomic DNA, liberating a 3' hydroxyl, which serves as a primer for reverse transcription of L1 RNA by L1 reverse transcriptase. Lastly, L1 cDNA joins genomic DNA, generating L1 structural hallmarks (i.e. mechanisms for frequent 5' truncations, a 3' A-tail, and variable length target site duplications) (Gilbert et al. 2005). The mechanisms for strand synthesis and the completion of L1 integration are unknown (Deininger et al. 2003). Most L1s are retrotransposition defective because they are 5' truncated, contain internal rearrangements, or harbor mutations within their open reading frames (Gilbert et al. 2005). SINEs are hypothesized to also utilize the TPRT method by hijacking the LINE proteins (Rinehart et al. 2005).

ID elements are believed to exploit LINE encoded proteins for retrotransposition. A retrotransposition assay will be used to investigate the utilization of LINE encoded proteins by ID elements in the process of retrotransposition. In future studies, plasmids containing ID elements from various rodents will be constructed with LINE containing plasmids and transfected into HeLa cells. Levels of retrotransposition can then be assessed by the surviving number of cells in the retrotransposition assay and correlated to the copy number levels within genomes of various rodents.

Methods and Materials

Various plasmids for different aspects of the retrotransposition assay were obtained from John Moran at the University of Michigan. Bacterial colonies were transformed using 100µl DH5α competent cells and 1µl plasmid stock; 30 min on ice, 45 sec at 42^{o}C, then 2 min on ice; 950µl SOC media was added and spread 20µl on an LB ampicillin (Amp) plate. The plates were then incubated at 37^{o}C overnight. A single

colony was selected and used to inoculate 50 ml liquid LB in an Erlenmeyer flask and grown at 37^{o} C. A large-scale isolation (Midiprep) of plasmid DNA was performed on 22 plasmids using the Promega Pure Yield™ Midiprep System Kit, according to the manufacturer's protocol.

DNA solutions were quantified using UV absorbance at 260 nm using the Beckman DU 640B spectrophotometer and the following equation: (260 nm reading) (OD_{260})(50µg/ml)(1ml/1000µl)(300µl volume) = µg DNA. These plasmids are to be used in future studies as variables and controls in a large scale HeLa cell retrotransposition assay. 0.5µg of the JM101L1.3 plasmid was digested with 5units each of the restriction enzymes *Not*1 and *Bst*Z11107I in a 20µl reaction. The pCEP vector band was separated from the LINE using a 0.8% agarose gel. The pCEP vector was cut out using a razor blade and was cleaned up using the Promega Wizard® Plus SV Gel and PCR Clean-up System, according to the manufacturer's protocol. DNA was quantified using the UV absorbance at 260 nm using the Beckman DU 640B spectrophotometer.

ID elements were isolated for mouse, rat, and deer mouse genomes using PCR to amplify the 5' end through the ID element of the BC1 gene (Figure 1) using the following conditions: In a 25µl volume, 2.5mM IDR-TR (GCCNAAATAAGTATACCCAGAGCTGAGGACCGA), 2.5mM BC1-F (AATGCGGCCGCATTTTGGAAGGTATCTCTGATG), 0.5µl DNA, and 1X Buffer (Epicentre Fail Safe Buffer D for mouse, rat, and deer mouse and Epicentre Fail Safe Buffer I for guinea pig and degu) using the Fail Safe™ PCR Premix Kit, according to the manufacturer's protocol. Thermocycler conditions were as follows: 94^{o}C: 2 min, (94^{o}C: 30 sec, 52^{o}C: 30 sec, 72^{o}C: 30 sec) repeat 32 times, 72^{o}C: 5 min, 4^{o}C. For guinea pig and

degu, the conditions were altered from a 52°C: 30 sec annealing step to a 55°C: 30 sec annealing step.

Clones of mouse, rat, and deer mouse PCR product of 5' through ID were generated by cloning, using the Promega pGem® -T Easy Vector System II, according to the manufacturer's protocol. A transformation of this ligation was conducted using 50µl JM109 competent cells and 2µl ligation reaction; iced for 20 min, 42°C for 45 sec, and iced for 2 min; 950µl of SOC was added and the solution was shaken at 140rpm for 1.5 hours at 37°C. 100µl of this solution was plated on an LB Amp plate that had been treated with 100µl IPTG and 20µl X-Gal. The plates were incubated overnight at 37°C. A small-scale isolation (Miniprep) of the plasmids of the observed colonies was done using the Wizard® Plus SV Miniprep: DNA Purification System, according to the manufacturer's protocol. The isolated plasmids were digested with two enzymes under the following conditions: 3µl plasmid, 5 units EcoR1, and 1X buffer in a 20 µl volume overnight at 37°C.

The 35µl of the original 5' through ID mouse, rat, and deer mouse PCR products were digested using 5units each of the restriction enzymes *Not*1 and *Bst*Z1107I, and a 1X buffer in a 50µl reaction. The restriction digested ID elements and JM101L1.3 vector were ligated together under the following conditions: 12ng digested JM101L1.3 vector, 10ng digested rodent PCR product, 1X buffer in a 10µl reaction at 14°C overnight using the Fermentas T4 DNA Ligation protocol. This ligation mix was used to transform 50µl JM109 competent cells by the Heirs Hock method; this involves placing on ice for 20 min, 42°C for 45 sec, and iced for 2 min; 950µl of SOC media were added and the solution was shaken at 140rpm for 1.5 hours at 37°C. 100µl of this solution was plated

on an LB Amp plate that had been treated with 100μl IPTG and 20μl X-Gal for blue/white colony selection.

Results

The ID elements from three different rodent genomes were amplified using PCR. These amplified SINEs were digested with restriction enzymes in preparation for incorporation into the pCEP vector of the JM101L1.3 plasmid. Before an incorporation of plasmid DNA into a HeLa cell retrotransposition assay can occur, the transformation of the ligation of a pCEP vector and rodent ID element must be isolated on a large scale.

Twenty-two 300μl plasmid DNA samples were isolated from bacterial stocks (provided by the John Moran Lab at the University of Michigan) in preparation for the HeLa cell retrotransposition assay. Concentrations of these plasmids were determined and recorded (Table 1). A 0.5ug JM101L1.3 sample was digested using the restriction enzymes *Not1 and BstZ1107I* in preparation for the incorporation of a rodent ID into the separated pCEP vector of the plasmid.

The digested JM101L1.3 plasmid was run on a 0.8% agarose gel to separate the pCEP vector from the LINE (Figure 2). Two DNA bands were observed on the gel. The larger band corresponded to the expected size of the pCEP vector (18,662bp) and the shorter band was comparable to the expected LINE size (6,500bp) (Figure 2). The pCEP vector for each of the four samples was isolated from the LINE on a 0.8% agarose gel, cleaned up using the Promega Wizard® Plus SV Gel and PCR Clean-up System and quantified. Each of the four samples was analyzed on a 0.8% agarose gel (Figure 2) and the concentrations of these vectors were recorded in μg/μl (Table 2).

The PCR amplified products of the mouse, rat and deer mouse ID was observed on a 2% agarose gel. The resultant fragments were approximately 300bp in length (Figure 3). This result corresponds to the expected 5' through ID length (Figure 1). The band observed in the negative control is not believed to be the result of contamination, but from the deer mouse sample flowing over into the negative control well.

The PCR products of guinea pig and degu were separated on a 2% agarose gel. Upon further applications of this technique, different patterns were observed on each agarose gel: either the gel did not contain a lone 300bp fragment and background fragments were present (Figure 4), or a 300bp fragment was observed with an ~300bp band in the negative control, strongly suggesting contamination (Figure 5). These results were not acceptable because they did not reliably isolate and amplify the ID of these two rodents.

The TA cloning of the mouse, rat, and deer mouse PCR amplification product resulted in three white mouse and two white deer mouse colonies on the LB Amp plates. A small-scale isolation followed by a restriction digest with EcoR1 revealed an insert of ~300bp (Figure 6).

The restriction digested PCR products of the mouse, rat and deer mouse ID elements were analyzed on a 2% agarose gel beside their undigested PCR counterparts, to ensure the digested fragment lengths were recovered. The digested fragment lengths were approximately 300bp (Figure 7).

The pCEP sample with the highest concentration (Table 2) was used for the ligation of mouse, rat and deer mouse IDs. The transformation and subsequent LB Amp

plating yielded one colony on each of the mouse and deer mouse plates and will be analyzed for potential use in a future retrotransposition assay.

Discussion

The retrotransposition of Class I TEs (SINEs and LINEs) occurs when reverse transcriptase of an RNA intermediate is used during the integration of mRNA into the genome (Deininger et al. 1992). The LINE sequence contains two ORFs that encode proteins that facilitate the retrotransposition of the mRNA. SINEs do not contain ORFs and require the enzymes encoded by autonomous TEs to ensure their transposition (Dewanniex et al. 2003). ID elements are thought to exploit the proteins LINEs during retrotransposition for their own integration into the genome. Therefore, ID elements and pCEP vectors were ligated and transformed in preparation into a HeLa cell retrotransposition assay.

The PCR and restriction digests of the mouse, rat, and deer mouse were successfully performed. The presence of a 300bp amplified DNA fragment for the mouse, rat, and deer mouse supports the successful amplification of the 5' through ID portion (Figure 1) of the rodent genome. The occurrence of product without the addition of DNA in the negative control of the guinea pig and degu PCR suggest contamination in the buffer and/or water. An improper annealing temperature could allow for other, less primer specific, regions of the rodent genome to amplify creating background bands when the 5' through ID DNA fragment was the intended target for amplification, particularly since ID is found in numerous copies within rodent genomes. Also, the guinea pig and degu sequences are more divergent from the mouse, rat, and deer mouse

sequences. This variation may include the primer-binding region, in which case a new set of primers, or a nested-PCR strategy, may prove effective.

The JM101L1.3 plasmid was successfully digested with restriction enzymes, separating the LINE from the pCEP vector. The presence of an 18,662bp and 6,500bp DNA fragments of the restriction digested JM101L1.3 plasmid concurs with the expected pCEP and LINE lengths respectively. The restriction digested rodent ID elements have been ligated into their respective plasmid vectors to generate the experimental constructs. This ligation was then transformed and bacterial colonies were generated. A large-scale isolation will be performed with the intention of incorporation into a future HeLa cell retrotransposition assay.

In the upcoming retrotransposition assay, experimental plasmid constructs containing the ID elements generated from mouse, rat, deer mouse, from the study as well as guinea pig and degu clones will be co-transfected with LINE containing plasmids in HeLa cells. Efficacy of ID integration will be determined via neomycin selection. Surviving HeLa cells will be analyzed for retrotransposition by establishing which SINEs are the best jumpers, correlating efficiency with genomic copy numbers, and determining if SINEs jump best with the LINEs of the same or different species.

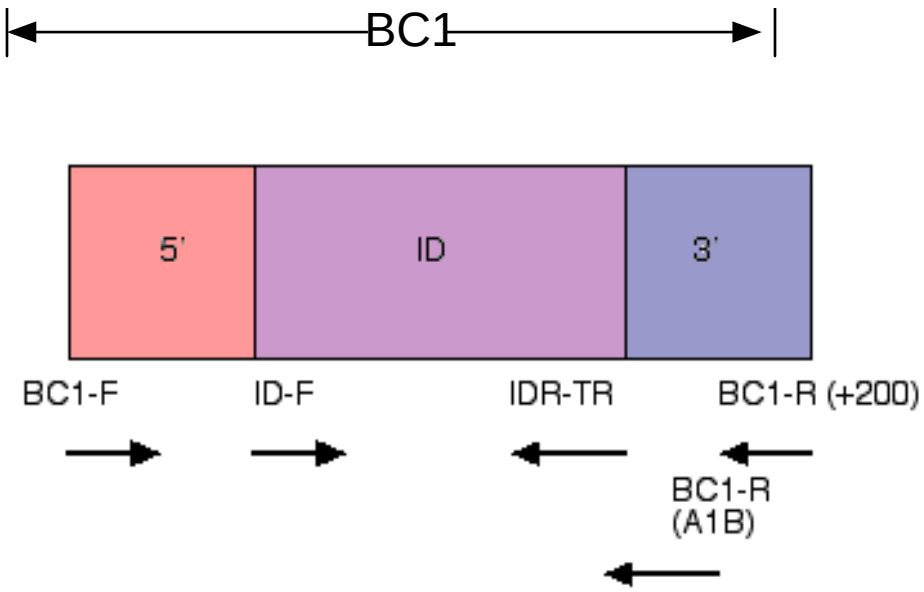

Figure 1: Primers used for amplification of different regions of the BC1 gene.

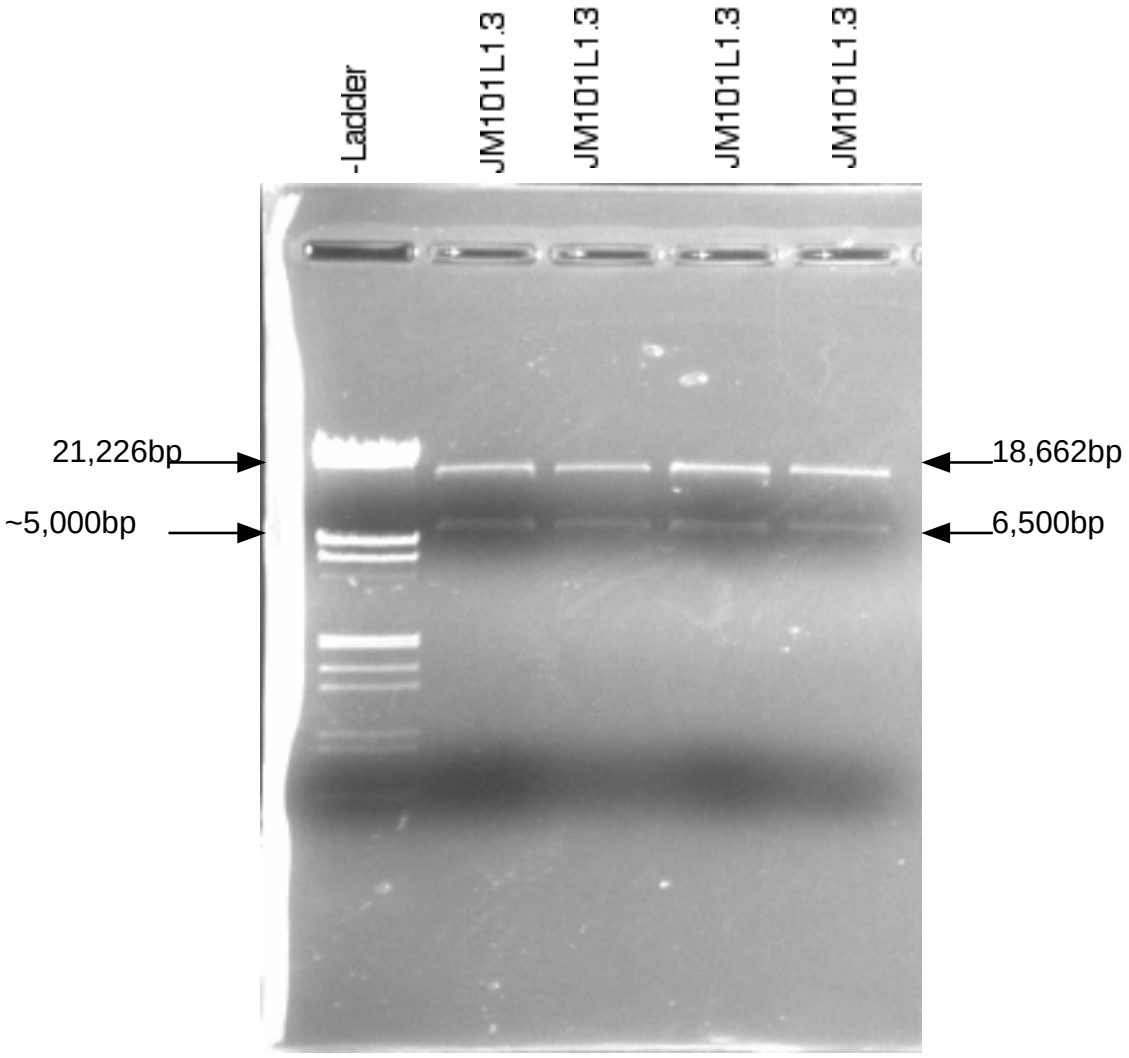

Figure 2. Restriction digest analysis of four large-scale plasmid preps of cloned JM101L1.3. LINE (6500bp) and vector (18,662bp) fragments were separated on a 0.8% agarose gel.

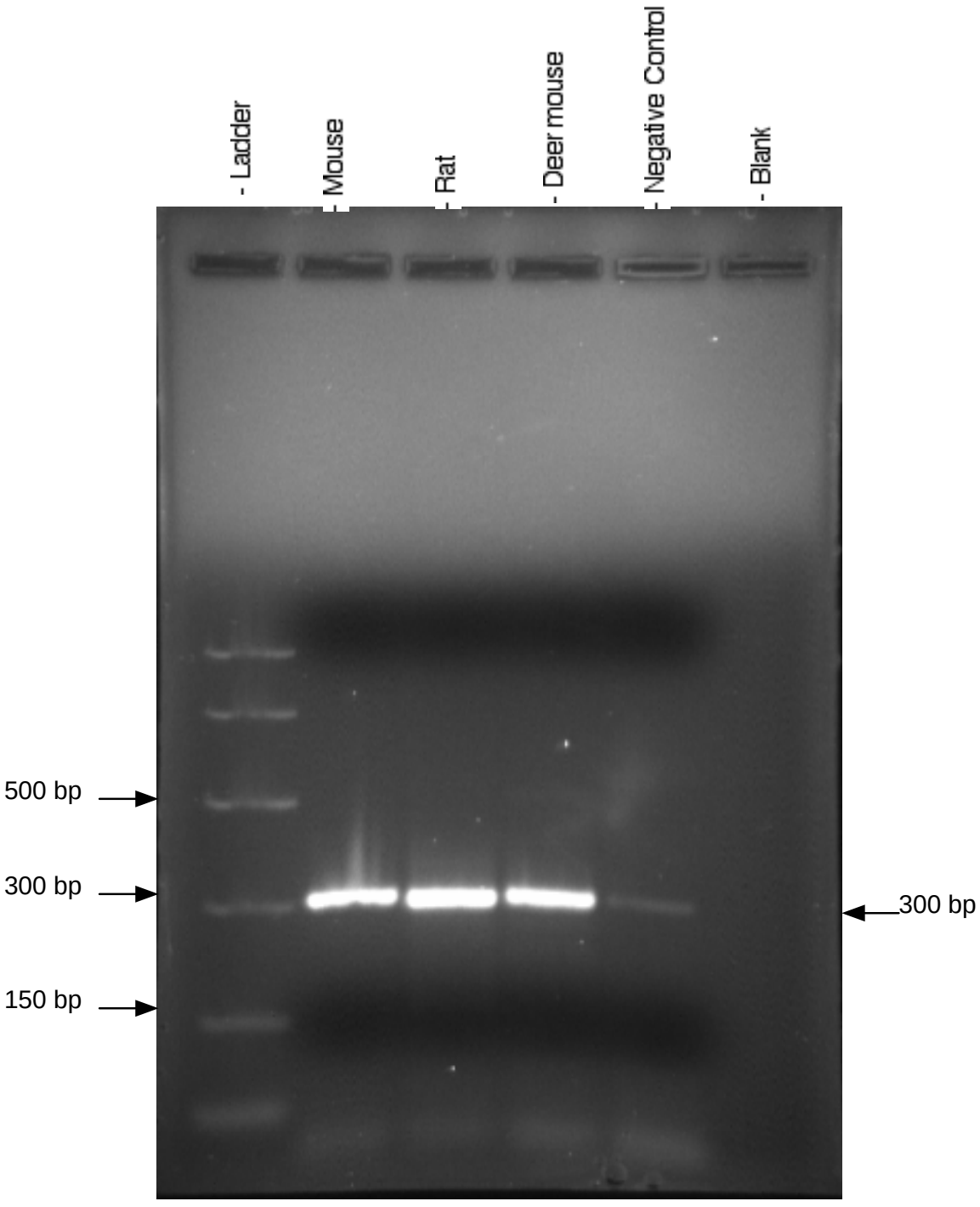

Figure 3. PCR amplified rodent ID elements analyzed on a 2% agarose gel.

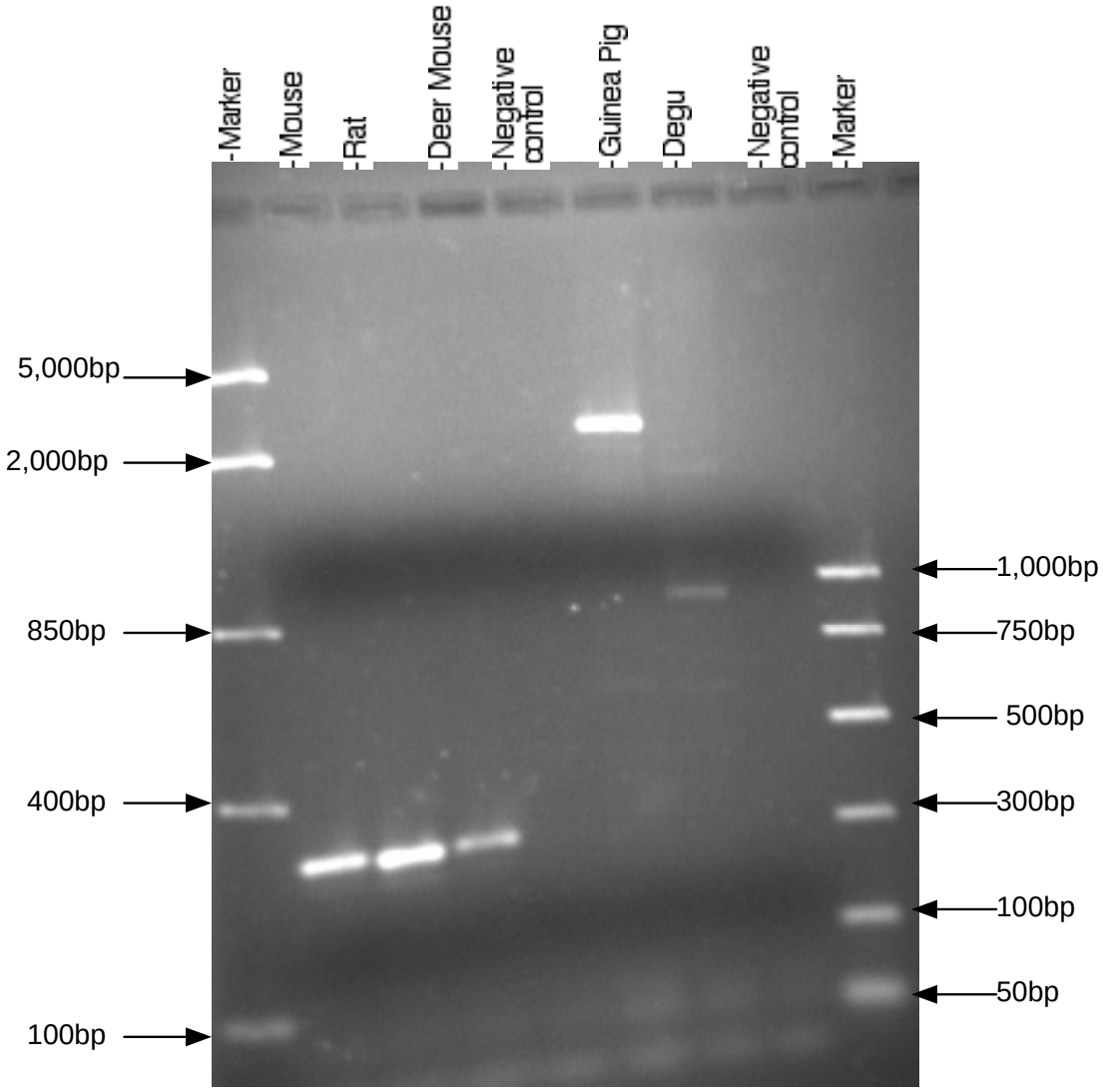

Figure 4. PCR amplified rodent ID elements analyzed on a 2% agarose gel. Mouse, rat, and deer mouse PCR products give an ~300bp band; guinea pig and degu each provide four bands ranging from ~3,000bp to ~600bp (fainter bands were observed on the original gel picture).

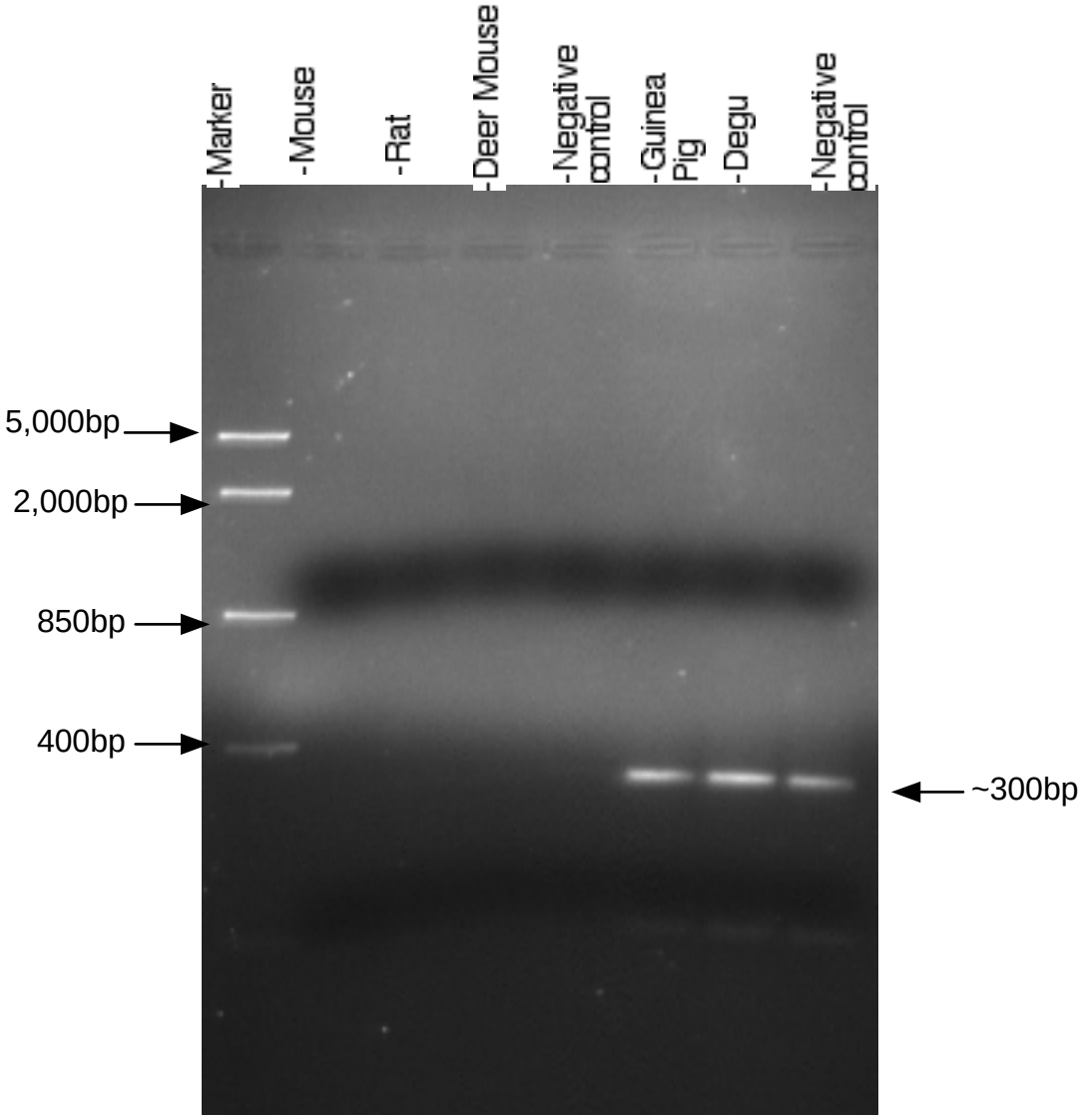

Figure 5. PCR amplified rodent ID elements analyzed on a 2% agarose gel. Mouse, rat, and deer mouse PCR did not amplify. The negative control for the guinea pig and degu PCR products has a ~300bp band.

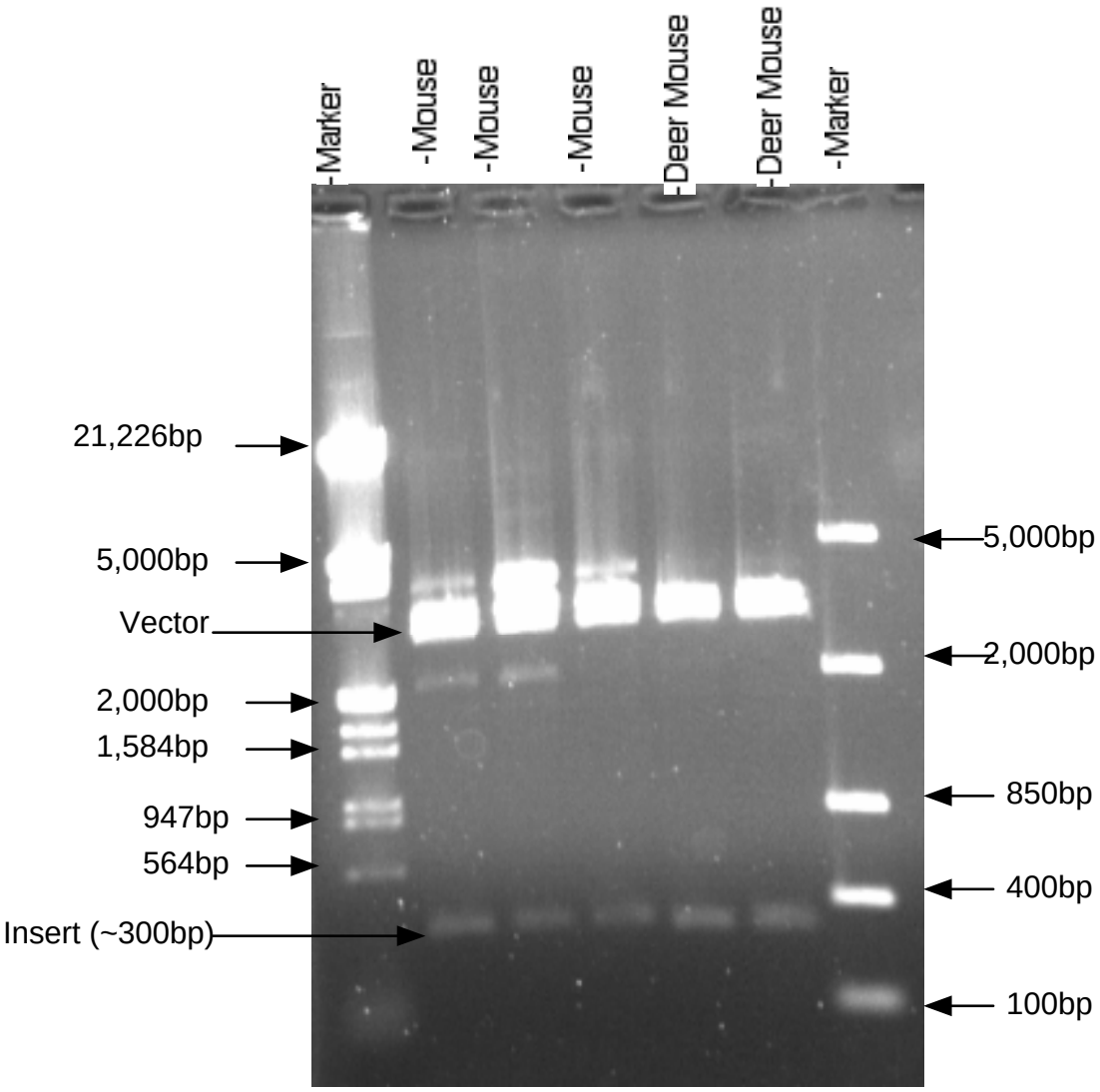

Figure 6. EcoR1 restriction digest of mouse and deer mouse TA clones. Samples exhibit the expected ~300bp insert on a 0.8% agarose gel.

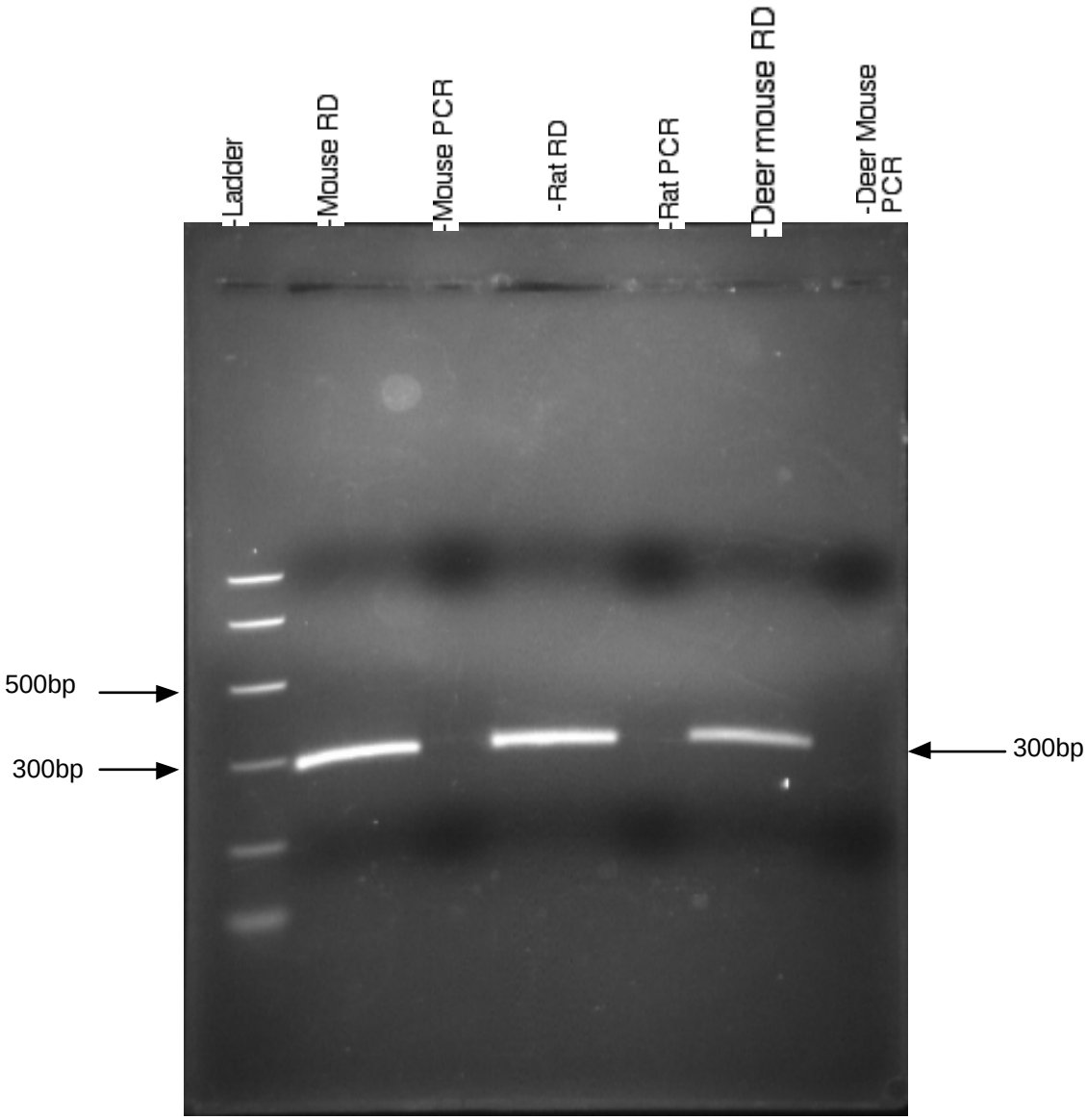

Figure 7. Restriction enzyme digestion of rodent ID PCR products is run beside their perspective PCR products on a 2% agarose gel to ensure accuracy.

Table 1. Isolation of various plasmid constructs using the Promega Pure Yield™ Midiprep System Kit protocol. Concentrations of plasmids are indicated.

Plasmid construct	Concentration
Alu Pure A	0.3561ug/ul
CEPB1	0.10125ug/ul
CEPΔ1933NF1	0.3ug/ul
CEPΔ1933NF1	0.0875ug/ul
CEPΔ1933SX3	0.1175ug/ul
CEPΔQ1933	0.0838ug/ul
CEPL1ddΔneo	0.556ug/ul
CEPL1NRPΔneo	0.03875ug/ul
CEPL1NRPΔneo	0.1837ug/ul
GFP	0.2375ug/ul
GFP	0.37375ug/ul
JM101	0.555ug/ul
JM105	0.2929ug/ul
JM101L1.3	0.2209ug/ul
JM101L1.3APRnoneo	0.7288ug/ul
KS1933	0.04127ug/ul
L1SMΔneo	0.6625ug/ul
ORF1Mneo	0.01625ug/ul
ORFMneo1	0.0405ug/ul
PCEPL1SM	0.14127ug/ul
PCEPL1SM	0.02375ug/ul
PBSNF1	0.505ug/ul

Table 2. pCEP vector concentrations in order as they appear from left to right in figure 2.

pCEP vector	Concentration
1	0.0065μg/μl
2	0.012μg/μl
3	0.0035μg/μl
4	0.007μg/μl

Bibliography

DeChiara TM, Brosius J. (1987). *Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content.* Proceedings National Academy of Science U S A. 84:2624-2628.

Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH. (1992). *Master genes in mammalian repetitive DNA amplification*. Trends in Genetics. 8:307-311.

Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. (2003). *Mobile elements and mammalian genome evolution.* Current Opinion of Genetics and Development. 13:651-658. Review.

Dewannieux M, Esnault C, Heidmann T. (2003). *LINE-mediated retrotransposition of marked Alu sequences*. Nature Genetics. 35:41-48.

Gilbert N, Lutz S, Morrish TA, and Moran JV. (2005). *Multiple Fates of L1 Retrotransposition Intermediates in Cultured Human Cells.* Journal of Molecular and Cellular Biology, 25: 7780-7795.

Kajikawa M, Ichiyanagi K, Tanaka N, and Okada N. (2005). *Isolation and Characterization of Active LINE and SINEs from the Eel*. Journal of Molecular Biology and Evolution 3:673-682.

Kass DH, and Batzer MA. (2003). *Transposable genetic elements*, in *Genetics* (Encyclopedia). R. Robinson (ed). New York: Macmillan Reference USA, Vol. 4:143-148.

Kass DH, Joomyeong K, Deininger PL. (1996). *Sporadic Amplification of ID Elements in Rodents*. Journal of Molecular Evolution. 42, 7-14.

Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P (1994). *Rodent BC1 RNA gene as a master gene for ID element amplification.* Proceedings National Academy of Science 91:3607-3611.

Kramerov DA, Vassetzky NS. (2005). *Short retroposons in eukaryotic genomes*. International Review of Cytology. 247:165-221.

Rinehart TA, Grahn RA, Wichman HA. (2005). *SINE extinction preceded LINE extinction in sigmodontine rodents: implications for retrotranspositional dynamics and mechanisms*. Cytogenet Genome Research. 110:416-425.

Vassetzky NS, Ten OA, Kramerov DA. (2003). *B1 and related SINEs in mammalian genomes*. Gene. 319:149-160.