Molecular Phylogenetics and Evolution 94 (2016) 518–530

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Emergence and evolution of *Zfp36l3* ☆

Timothy J. Gingerich [a], Deborah J. Stumpo [a], Wi S. Lai [a], Thomas A. Randall [b], Scott J. Steppan [c], Perry J. Blackshear [a,d,e,*]

[a] Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA
[b] Integrative Bioinformatics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA
[c] Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA
[d] Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA
[e] Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

ARTICLE INFO

Article history:
Received 26 June 2015
Revised 6 October 2015
Accepted 13 October 2015
Available online 19 October 2015

Keywords:
Gene evolution
New gene formation
Retrotransposon
Intronization
Repetitive element expansion
Subcellular localization

ABSTRACT

In most mammals, the *Zfp36* gene family consists of three conserved members, with a fourth member, *Zfp36l3*, present only in rodents. The ZFP36 proteins regulate post-transcriptional gene expression at the level of mRNA stability in organisms from humans to yeasts, and appear to be expressed in all major groups of eukaryotes. In *Mus musculus*, *Zfp36l3* expression is limited to the placenta and yolk sac, and is important for overall fecundity. We sequenced the *Zfp36l3* gene from more than 20 representative species, from members of the Muridae, Cricetidae and Nesomyidae families. *Zfp36l3* was not present in Dipodidae, or any families that branched earlier, indicating that this gene is exclusive to the Muroidea superfamily. We provide evidence that *Zfp36l3* arose by retrotransposition of an mRNA encoded by a related gene, *Zfp36l2* into an ancestral rodent X chromosome. *Zfp36l3* has evolved rapidly since its origin, and numerous modifications have developed, including variations in start codon utilization, *de novo* intron formation by mechanisms including a nested retrotransposition, and the insertion of distinct repetitive regions. One of these repeat regions, a long alanine rich-sequence, is responsible for the full-time cytoplasmic localization of *Mus musculus* ZFP36L3. In contrast, this repeat sequence is lacking in *Peromyscus maniculatus* ZFP36L3, and this protein contains a novel nuclear export sequence that controls shuttling between the nucleus and cytosol. *Zfp36l3* is an example of a recently acquired, rapidly evolving gene, and its various orthologues illustrate several different mechanisms by which new genes emerge and evolve.

Published by Elsevier Inc.

1. Introduction

Zinc finger protein 36-like 3 (ZFP36L3) belongs to a small family of RNA binding proteins, known as the tristetraprolin or TTP family, that regulate gene expression by binding to AU-rich elements within certain mRNAs and accelerating their rates of degradation (Blackshear, 2002; Brooks and Blackshear, 2013; Carrick et al., 2004). The other mammalian family member proteins are ZFP36 (also known as tristetraprolin or TTP), ZFP36L1 (also known as cMG1, ERF1 and BRF1) and ZFP36L2 (also known as ERF2 and BRF2). The defining characteristics of members of this protein family in eukaryotes are a central tandem zinc finger (TZF) domain, consisting of two highly conserved zinc fingers of the CCCH class, and a C-terminal sequence that is thought to associate with the NOT1 scaffolding protein (Blackshear and Perera, 2014; Fabian et al., 2013). Although all mammalian family members exhibit similar mRNA binding and mRNA destabilizing properties in cultured cells (Blackshear et al., 2005; Lai et al., 2000), they are differentially expressed in tissues and in response to stimuli, suggesting that they can control the stability of different sets of mRNAs under normal physiological circumstances. This is supported by KO mice for the genes encoding the first three family members, *Zfp36*, *Zfp36l1* and *Zfp36l2*, all of which have very different phenotypes (Stumpo et al., 2009, 2004; Taylor et al., 1996).

Direct orthologues of TTP, ZFP36L1 and ZFP36L2 have been found in essentially all vertebrate species examined, with the exception of birds, which appear to lack TTP (Lai et al., 2013b). Proteins with conserved TZF domains and NOT1 binding domains that otherwise bear little resemblance to the mammalian ZFP36 family members are found in most groups of eukaryotes, including fungi and plants, indicating that these genes have ancestral roots dating

☆ This paper was edited by the Associate Editor Derek Wildman.

* Corresponding author at: 111 TW Alexander Drive, Research Triangle Park, NC 27709, USA. Fax: +1 (919) 541 4571.
E-mail address: Black009@niehs.nih.gov (P.J. Blackshear).

http://dx.doi.org/10.1016/j.ympev.2015.10.016
1055-7903/Published by Elsevier Inc.

back more than a billion years (Blackshear and Perera, 2014; Lai et al., 2013b).

Previous studies from our laboratory described the initial identification and characterization of mouse ZFP36L3 (Blackshear et al., 2005; Frederick et al., 2008). Although this protein behaved like other TTP family members in its ability to promote mRNA deadenylation and decay, it differed from the other family members in several ways, including: (1) a much larger size and an acidic isoelectric point; (2) the absence of an intron in the protein coding region; (3) the lack of nucleo-cytoplasmic shuttling, apparently because of a novel, long, C-terminal alanine-rich repeat domain that maintains the cytosolic localization of the protein; (4) the presence of a proline residue in place of an ancestral leucine in the lead-in sequence to the second zinc finger; (5) the presence of an acidic residue in place of one of the conserved hydrophobic residues within a C-terminal domain, that destroys a functional nuclear export sequence; (6) expression only in the placenta and yolk sac, whereas the other family members are widely expressed in mouse tissues; and (7) the apparent lack of an orthologue in the human and other non-rodent mammalian genomes (Blackshear et al., 2005; Frederick et al., 2008). Within the placenta, it was highly expressed within the syncytiotrophoblast cells of the labyrinthine zone, suggesting that it might be important for the physiological function of the placenta in mice. Recent studies of a *Zfp36l3* KO mouse have demonstrated that ZFP36L3 is important for overall fecundity in this species, and regulates the placental expression of many genes important for normal placental physiology (D.J. Stumpo and P.J. Blackshear, manuscript submitted).

In this study, our main goal was to examine the phylogenetic distribution of ZFP36L3, and, from that information and current knowledge of rodent evolution, estimate the evolutionary period when it first arose, and the mechanism of its formation. A secondary goal was to explore the evolutionary diversity of this protein as it relates to its function, particularly since it differs in many respects from other members of its protein family. We searched the available sequenced species and determined that *Zfp36l3* was not present in Dipodidae, or any earlier branches of the rodent order. We concluded that the *Zfp36l3* gene arose in an ancestor of the Muridae, Cricetidae and Nesomyidae families, with the absent expression in the Spalacidae defining the evolutionary boundary of the expressed gene. The origin of *Zfp36l3* appears to have been a retrotransposition event involving the mRNA encoded by its related family member, *Zfp36l2*. Since its first appearance, it has accumulated numerous genetic modifications among the descendant species. Some of these provide insights into the mechanisms by which new genes originate and how they subsequently evolve, while others provide insights into the function of this interesting family of RNA binding proteins.

2. Materials and methods

2.1. Acquisition of new sequence data

All newly generated sequences were submitted to Genbank; their accession numbers are provided in Tables S1 and S2. *Zfp36l3* and other sequences from *Cricetomys gambianus* were generated using the Illumina Nextseq 500 platform. The DNA was prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina), as recommended. Note that DNA was sheared to an average size of 1.5 kb using a Covaris S220 focused-ultrasonicator with the following settings: 50 s treatment time, 175.0 watts peak power, 5% duty factor, and 200 cycles/burst. The NextSeq 500, High Output, 300 cycle Kit was used to generate a total of 373,179,044 paired end reads from a single library. QC of these data was done with fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). These data were imported into CLC Genomics 7.0.3 (Qiagen, Aarhus, Netherlands) running on a Dell T7610. A *de novo* sequence assembly was performed (word size = 64, bubble size = 80; all other settings were default). An assembly of 521,648 contigs covering 2.18 Gb of the genome was generated with a contig N50 of 9.2 kb, and a scaffold N50 of 36.4 kb. RefSeq mRNA and protein sequences of genes of interest from mouse (mm10) and rat (rn5) were used to BLAST a database of these contigs, and the individual contigs containing complete genes were extracted for further analysis. The sequence data from this experiment have been deposited in the GenBank Single Read Archive as SRP059556.

Other *Zfp36l3* sequences were amplified from genomic DNA samples by PCR, using Platinum Taq DNA Polymerase High Fidelity (Life Technologies). PCR products were gel purified using the QIAquick gel extraction kit and the fragments were cloned into the sequencing plasmid pCR2.1-TOPO, using the TOPO TA cloning kit (Life Technologies). DNA was sequenced by Genewiz, (Durham, NC), using M13 and T7 primer sites present in the TOPO plasmid. In cases where cloned fragments were too long to read through from their ends, inserts were sequenced by primer walking. In many cases, degenerate primers were designed based on related species, and the nucleotide inosine was substituted at positions where there was uncertainty between thymine and guanine bases. Sequence reads were assembled using the CAP3 contig assembly program, available online: http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::cap3.

Sequences for the 5′ UTRs of *Mesocricetus auratus*, *Meriones unguiculatus* and the *Peromyscus* species *maniculatus* and *polionotus* were acquired by 5′ RACE, using the protocol provided in the 5′ RACE System for Rapid Amplification of cDNA Ends, version 2.0 kit (Life Technologies). Sequences for the 3′ UTRs of *Rattus norvegicus*, *Cricetulus griseus* and *Mesocricetus auratus* were acquired using the protocol provided in the 3′ RACE System for Rapid Amplification of cDNA Ends kit (Life Technologies).

RNA was purified from cell lines and tissues using Trizol (Life Technologies). Reverse transcription was performed using the SuperScript III reverse transcription kit (Life Technologies). Genomic DNA was isolated from tissues using proteinase K digestion, followed ethanol precipitation, as described (Laird et al., 1991).

2.2. DNA and tissue sources

Sources of genomic DNA and tissues are listed in Table S4.

2.3. Northern blotting

Northern blots were prepared and hybridized to random primed, ^{32}P dCTP labeled probes as described previously (Frederick et al., 2008). The probe used for detecting the *Mus musculus* and *Rattus norvegicus* Zfp36l3 transcripts was generated using template DNA encompassing the full open reading frame of *Mus musculus* Zfp36l3. The probe for detection of *Sigmodon hispidus* Zfp36l3 mRNA was generated using DNA spanning the entire open reading frame of the transcript from this species. The probe used to detect the *Peromyscus maniculatus* and *Peromyscus polionotus* transcripts encompassed 637 bases of the *Peromyscus maniculatus* transcript, starting in the 5′ UTR and ending 388 bases after the start codon. The probe for the *Cricetulus griseus* transcript targeted 1046 bases of the open reading frame, beginning 459 bases downstream of the start codon.

2.4. Plasmids

Plasmids expressing *Mus musculus* and *Peromyscus maniculatus* ZFP36L3, with N- and C-terminal fusions to EGFP, were constructed

by PCR amplification of these open reading frames from placental cDNA, followed by restriction digestion and ligation into the EGFP vectors as described below. Primers used for cloning the open reading frames are listed in Table S5. The plasmids with EGFP fused to the C-terminus of ZFP36L3 were constructed by digesting the inserts and the vector CMV.EGFP.BGH3′/BS+, described previously (Lai et al., 2013a), with HindIII and BamHI, followed by ligation. The plasmids with EGFP fused to the N-terminus of ZFP36L3 were constructed by digesting the inserts and the vector, CMV2.EGFP(c).BGH3′/BS+ (described below), with BamHI and NsiI, followed by ligation. Plasmid CMV.EGFP(c).BGH3′/BS+ was constructed as follows: A fragment encoding the EGFP protein (without the stop codon) was amplified by PCR from pEGFP-N1 (Clontech), using forward and reverse primers containing HindIII and Asp718 sites, respectively. This EGFP fragment, and plasmid CMV.BGH3′/BS+, described previously (Lai et al., 1999), were cut with HindIII and Asp718, and ligated to create CMV.EGFP(c).BGH3′/BS+.

The plasmid expressing untagged *Peromyscus maniculatus* ZFP36L3 was generated by amplifying the ORF from placental cDNA by PCR, using primers listed in Table S5. The plasmid CMV.EGFP(c).BGH3′/BS+, was cut with HindIII and XhoI to remove EGFP, and was then ligated with the *Peromyscus maniculatus* ZFP36L3 ORF, cut with the same enzymes.

NetNES 1.1 identified two possible NESs around amino acids 10–13 and 298–304 of the *Peromyscus maniculatus* ZFP36L3 protein. These sites were deleted in the plasmids expressing the N- and C-terminal *Peromyscus maniculatus* ZFP36L3-EGFP fusions using the protocol associated with the QuikChange II site directed mutagenesis kit (Agilent Technologies). PCR was performed using complementary primers that contained the desired deletions (Table S5) and Phusion High-Fidelity DNA Polymerase (NEB). The PCR parameters were: denaturation at 98 °C for 30 s, followed by 18 cycles of 98 °C for 9 s, 55 °C for 20 s and 72 °C for 4 min. Following PCR, 1 μL of 20 units/μL DpnI (NEB) was added and the samples were incubated at 37 °C for 1 h. A 5 μL aliquot of this was transformed into One Shot TOP10 Competent Cells (Life Technologies), and plasmids were prepared and deletions were confirmed by sequencing.

Plasmids for bacterial expression of poly-histidine (His_6) tagged *Peromyscus maniculatus* ZFP36L3 fragments (amino acids 1–130 and 198–350) were constructed as follows: ZFP36L3 encoding fragments were generated by PCR (see Table S5 for primers) and ligated into the donor vector pENTR/TEV/D-TOPO (Life Technologies). The inserts were then transferred to the Gateway destination vector pDEST-527-Tev (kindly provided by Dr. Dominic Esposito, NCI/SAIC/Maryland).

2.5. Cell culture and transfections

HEK 293 cells were maintained in DMEM, supplemented with 10% (v/v) fetal bovine serum. Cells were plated onto 35 mm glass bottom plates (MatTek Corp) and grown to roughly 50% confluence before transfecting. Transient transfections were performed using 5 μL of Lipofectamine 2000 and 100 ng of plasmid. Following transfection, the cells were grown for 24 h, with one media change after 8 h. For some experiments, cells were treated with 10 ng/mL Leptomycin B (Sigma) or carrier (methanol:water (7:3)) for 3 h prior to fixation.

2.6. Immunostaining and imaging

For GFP imaging, transfected cells were fixed in glass bottom plates, in a solution of 4% paraformaldehyde in PBS for 20 min, then quenched in a solution of 0.1 M glycine in PBS for 15 min, with PBS washes between each step. After the last wash was removed, the cells were mounted using two drops of ProLong Gold antifade reagent (Life Technologies) and covered with an 18 mm, round coverslip (neuVitro). For immunostaining, cells were fixed with 4% paraformaldehyde in PBS for 5 min, quenched with 0.1 M glycine for 5 min, fixed with ice cold methanol for 5 min and permeabilized with 0.2% Triton-X-100 for 20 min, with PBS washes between steps. The cells were then blocked with 3% BSA for 1–2 h. Rabbit anti-*Peromyscus maniculatus* ZFP36L3 antibody (#2050) was diluted 1/500 in 0.3% BSA, and incubated with the cells overnight at 4 °C. Goat anti-rabbit secondary antibodies, conjugated to Alexa 488 or Alexa 594 (Life Technologies), were diluted 1/1000 in 0.3% BSA, and left on the cells for approximately 2 h, followed by 4 washes. Note that in the steps following incubation with antibodies, all washes were performed using 0.3% BSA in PBS and that all solutions used in the fixation, immunostaining and washing steps were made with PBS and performed in the MatTek plates, using 1 mL volumes. Images were obtained using a Zeiss LSM710 confocal microscope, with the 40× oil immersion lens.

2.7. Antibody production

Bacterial expression plasmids (described in Section 2.4) were generated that encoded poly-histidine (His_6) fusions to the N-terminal ends of *Peromyscus maniculatus* ZFP36L3 peptides (amino acids 1–130, and amino acids 198–350). Note that the highly conserved zinc finger region is situated between the two fragments, and was omitted in order to reduce the potential for generating antibodies that cross-reacted with the other family members. The plasmids were transformed into Rosetta 2 DE3 pLacI cells and expression was screened at 18 °C and 30 °C. Both peptides were insoluble, but expressed well and accumulated in inclusion bodies. Scaled up expression was performed in 1 L of LB supplemented with 100 μg/mL ampicillin and 25 μg/mL chloramphenicol. Cultures were grown overnight at 37 °C with 220 rpm shaking, pelleted and resuspended in 50 mL of LB. That culture was split and used to inoculate two 1 L LB/amp cultures grown at 37 °C for 30 min, with 220 rpm shaking. Cultures were then induced with 1 mM IPTG for 3 h, pelleted and frozen at −80 °C. Inclusion bodies were purified with BugBuster master mix (Millipore), following the reagent protocol. Inclusion body pellets were resuspended in 50 mM HEPEs pH 7.5, with 6 M urea, and placed on a nutator overnight at 4 °C. The solution was centrifuged at 20,000*g* to remove insoluble-flocculent material, and the soluble fraction was then purified by binding to NINTA. Washes were performed with 20 column volumes of the following buffer: 50 mM HEPES, pH 7.5, 350 mM NaCl, and 6 M urea. The peptides were then eluted with 50 mM HEPES, pH 7.5, 250 mM NaCl, 6 M urea, and 250 mM imidazole. Eluates were resolved on a 12% Bis-Tris SDS gel (Invitrogen) and stained with Coomassie blue. The bands containing the peptides were cut from the gel and ZFP36L3 identities were confirmed by mass spectrometry.

Antibodies were generated by Harlan Bioproducts for Science (Indianapolis, IN). The gel slices containing the peptides were pulverized and approximately 500 μg of each peptide was injected into two rabbits. Rabbits 2050 and 2051 received the peptide comprising the N-terminal end of the *Peromyscus maniculatus* ZFP36L3 protein (amino acids 1–130), while rabbits 2052 and 2053 received the C-terminal peptide (amino acids 198–350). Following the primary immunization, rabbits were boosted a total of 4 times to increase titers. ELISA assays were performed by Harlan before the final bleeds and indicated the presence of antibodies to the ZFP36L3 peptides. The four antibodies were tested by Western blotting, and three (2050, 2051 and 2053) were reactive to their respective immunizing peptides, and to protein lysates from cells transfected with plasmids expressing *P. maniculatus* ZFP36L3, with little background (data not shown). Antibody 2052 showed high background and was not used.

2.8. Phylogenetic reconstruction of Zfp36l3 evolution

The presence or absence of *Zfp36l3* was mapped using maximum parsimony in Mesquite v. 2.5 (http://mesquiteproject.wikispaces.com/) onto a time-calibrated phylogeny of myodont rodents. This tree was derived from the 300-species chronogram of Schenk et al. (2013), and pruned down to the species sampled here. That analysis included 13 fossil calibrations to conduct a relaxed-molecular clock Bayesian analysis. Additional outgroup species were grafted onto the chronogram based on broader rodent phylogenetic studies (Fabre et al., 2012; Huchon et al., 2007), with dates scaled to the Schenk et al. timescale (that estimated dates approximately 15% younger than by the other two studies).

3. Results

3.1. Origin of Zfp36l3

Proteins with the characteristics of TTP family members are found in species as disparate as mammals, protists and plants (Blackshear and Perera, 2014). Only a single family member is expressed in non-chordate deuterostomes, such as sea urchins and sea squirts. The genes encoding the single TTP family members present in both species contain a single intron, and the sites of splicing are 20 and 21 amino acids from the N-terminus for the sea urchin *Strongylocentrotus purpuratus* (XP_782811.1) and the urochordate *Ciona intestinalis* (NP_001071879.1), respectively. The presence of a single intron, and its placement near the N-terminus of the proteins, has persisted through subsequent genome duplication events in the evolution of the vertebrate proteins, so that the three widely expressed family members in mammals (*Zfp36*, *Zfp36l1* and *Zfp36l2*) all contain a single intron at that site. When we initiated these studies, mRNA sequences for *Zfp36l3* were available for two murine species, *Mus musculus* and *Rattus norvegicus* (Blackshear et al., 2005). However, alignment of the *Mus musculus* and *Rattus norvegicus* Zfp36l3 mRNA sequences to the reference genomes of both species indicated that no such intron was present in the *Zfp36l3* genes from either species (Blackshear et al., 2005). This observation strongly suggests that the *Zfp36l3* genes in these species were derived from a related spliced mRNA through a process of retrotransposition, as discussed here, rather than evolving from a single invertebrate precursor.

When the encoded protein sequences from mouse and rat ZFP36L3 were compared to the other three gene-family members in each species, the greatest similarity was to ZFP36L2 in both cases. Specifically, a simple blastp comparison of *Mus musculus* ZFP36L3 with other family members from the same species yielded similarity scores of 1e-42 with ZFP36L2, 1e-36 with ZFP36L1, and 4e-26 with TTP. A similar comparison of *Rattus norvegicus* ZFP36L3 with the other family members from the same species yielded similarity scores of 4e-44 with ZFP36L2, 1e-38 with ZFP36L1, and 2e-28 with TTP. These results strongly suggest that an ancestral Zfp36l2 mRNA sequence was the source of the original *Zfp36l3* gene through a process of retrotransposition.

3.2. Phylogenetic distribution of Zfp36l3

Northern blotting of total placental RNA demonstrated bands hybridizing to a mouse Zfp36l3 cDNA probe in rat, golden hamster (*Mesocricetus auratus*), and gerbil (*Meriones unguiculatus*), but not in guinea pig (*Cavia porcellus*) (data not shown). Database searches and northern blotting repeatedly failed to identify related sequences in non-rodents, including humans. More recently (11/5/13), we searched genomic sequences from a variety of rodents that have been categorized by the NCBI as having "full genome representation", including *Cavia porcellus*, *Ictidomys tridecemlineatus* (13-lined ground squirrel), *Chinchilla lanigera* (long-tailed chinchilla), *Heterocephalus glaber* (naked mole rat), *Jaculus jaculus* (lesser Egyptian jerboa), *Octodon degus* (degu) and *Dipodomys ordii* (Ord's kangaroo rat). No orthologues of *Zfp36l3* were identified in any of these species. Very recently, genomic sequences and placental mRNA sequence reads became available for the Upper Galilee mountains blind mole rat, *Nannospalax galili*. Transcript sequences have been predicted in this species that encode proteins that are clearly orthologues of *M. musculus* TTP (XP_008852182.1; 8e-143 compared to *M. musculus* TTP), two isoforms of ZFP36L1 that differ by 23 amino acids at their extreme N-termini (XP_008850391.1 and XP_008850392.1; both e0.0 compared to *M. musculus* ZFP36L1), and ZFP36L2 (XP_008834094.1; e0.0 compared to *M. musculus* ZFP36L2). A fourth potential transcript was also identified, encoding a putative protein that bears some relationship to *Mus musculus* ZFP36L3 (XP_008828625.1; top *M. musculus* hit is ZFP36L3, 7e-44). An alignment of this putative protein with *M. musculus* ZFP36L3 is shown in Fig. S1. This alignment shows regions of sequence similarity in the TZF domain and near the C-terminus; however, the "diagnostic" proline residue leading into the second zinc finger and the aspartate residue four amino acids from the C-terminus were not present (Fig. S1). This potential transcript apparently derives from an intronless gene, and is thus likely to be the product of a retrotransposition event, again from the Zfp36l2 mRNA. However, an analysis of mRNASeq reads from this and a related species, *Spalax carmeli* (GenBank SRA number SRX1023526), found no evidence of mRNA expression of this fourth potential family member in either the placenta or brain, despite abundant expression of the other three transcripts in this gene family. We concluded from this analysis that the identified sequence is most likely a non-expressed pseudogene in these species. It was not possible to determine whether this presumed pseudogene is on the X chromosome in *N. gallili*, since the contig containing the probable pseudogene did not align with any particular *M. musculus* chromosome. One possibility is that this apparent pseudogene derives from an independent insertion of Zfp36l2 mRNA sequences into a new chromosomal locus in this species. Alternatively, it could be the product of the same retrotransposon insertion that led to the emergence of *Zfp36l3* as an expressed gene, but either lost promoter function, or branched evolutionarily before it gained promoter function; it also did not gain many of the other amino acid changes characteristic of ZFP36L3. Whatever the origin of this sequence, it appears that the lack of expressed Zfp36l3 mRNA in the Spalacidae define an evolutionary boundary for the modern species that express ZFP36L3.

Within the families Muridae, Cricetidae and Nesomyidae, sequences were acquired for the open reading frame of *Zfp36l3*, and in some cases the flanking regions and transcripts, from a variety of sources, including sequence repositories, direct sequencing of cloned genomic DNA and, where possible, placenta-derived cDNA. Table S1 lists, and Fig. 1 illustrates, the species for which *Zfp36l3* sequence data currently exist, along with species that were determined to be negative for *Zfp36l3*. All genomic and mRNA sequences that we generated for this study have been deposited in GenBank; accession numbers are listed in Tables S1 and S2. An alignment of all available full-length ZFP36L3 protein sequences is shown in Fig. S2, along with available sequences of ZFP36L2 from selected other rodents, including some that do not express ZFP36L3.

We mapped the presence and absence of *Zfp36l3* using maximum parsimony onto a pruned time-calibrated phylogeny derived from Schenk et al. (2013) (Fig. 1). The reconstruction is unequivocal, with *Zfp36l3* originating in an ancestor common to Muridae, Cricetidae and Nesomyidae (each other's closest relatives and members of Muroidea) between 23.7 and 37.8 million years ago. The approximate timing of subsequent evolutionary modifications to the protein sequences will be discussed in the sections below.

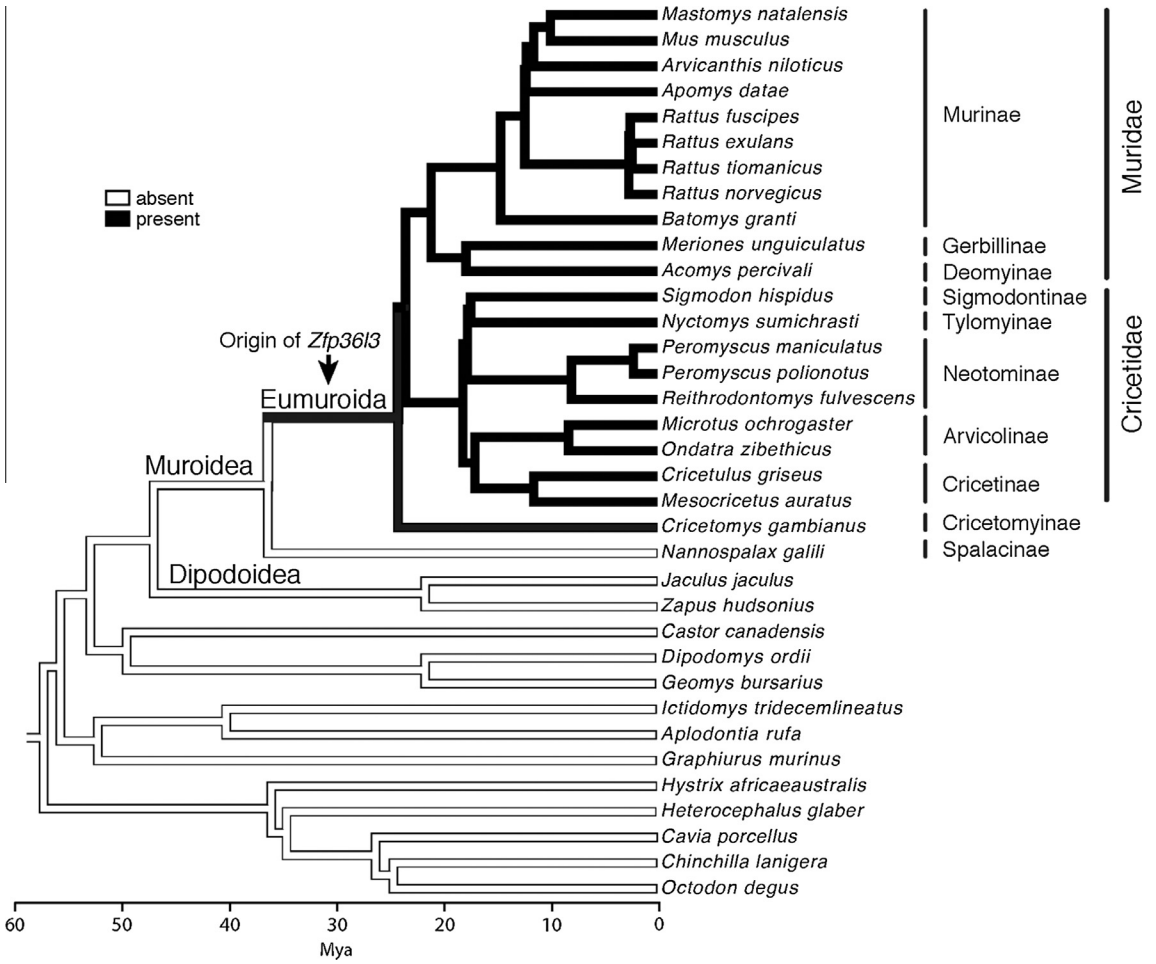

Fig. 1. Phylogenetic tree showing the evolutionary history of *Zfp36l3*. The presence of the gene was mapped using parsimony on a time-calibrated chronogram pruned from the 300-species phylogeny of Schenk et al. (2013). Black branches indicate presence of *Zfp36l3*, and white branches absence. The arrow points to the branch in the tree on which *Zfp36l3* is reconstructed to have evolved, between 23.7 and 37.8 million years ago.

3.3. Gain of an intron in Mus musculus Zfp36l3

An alignment of the full-length *Mus musculus* Zfp36l3 mRNA (NM_001009549.2) with the *Rattus norvegicus* mRNA (BK008909) revealed an unexpected gap in the alignment within the 3′ UTR. Specifically, the alignment was lost after b 2327 in the mouse mRNA and b 2169 in the rat mRNA. Some of the mouse sequence 3′ of this loss of alignment was very similar to sequences from mouse and rat ribosomal protein L37 (*Rpl37*) (NM_026069.3 for mouse: similarity score of 4e-45 over 145 b of alignment; NM_031106.1 for rat: similarity score of 7e-53 over 145 b of alignment). There was also an unexpected gap when the full-length mouse mRNA was aligned with the reference mouse genome, in that there was an apparent intron at approximately the same region as the *Rpl37* sequences. This intron, comprising the reverse complement of b 53772954-53774048 of NC_000086.7, contained sequences that were highly related to the mouse Rpl37 mRNA sequence (similarity score 9e-76).

A comparison between the rat mRNA and rat reference genome revealed that there was no evidence of either insertion of *Rpl37* sequences, or splicing out of an intron, in the analogous site within the 3′ UTR. Specifically, the rat mRNA is identical to the genomic sequence along its entire length, including its 1110 b 3′ UTR, and thus is an example of an intronless gene. The rat 3′ UTR exhibits some similarity to that of the rat Zfp36l2 mRNA (similarity score 2e-10 over 126 b), again supporting Zfp36l2 mRNA as the origin of *Zfp36l3*. The predicted size of the full-length rat Zfp36l3 mRNA, excluding the poly(A) tail, is 3330 b, which we have deposited in GenBank as accession number BK008909; this is in contrast to the mouse mRNA size of 2613 b, excluding the poly(A) tail. To confirm that these mRNA sequences correspond to the actual size of the transcripts in placenta, we performed northern blots of total cellular RNA from this tissue (Fig. 2A). The much smaller size of the mouse mRNA confirms that splicing of an intron has occurred in the mouse mRNA, resulting in a transcript that is approximately 0.7 kb smaller than the rat transcript.

The *Rpl37*-like DNA sequence that is inserted into the *Mus musculus Zfp36l3* gene is flanked by direct repeats and a poly A tail remnant (Fig. 2B) and provides a classic example of a retrogene, also referred to as a processed pseudogene (Vanin, 1985). Note that this presumed pseudogene is in reverse orientation with respect to the *Zfp36l3* gene. Although the precise mechanism by which processed pseudogenes are formed is not well documented, a very plausible model has been proposed (Kaessmann et al., 2009). Processed pseudogenes are believed to derive from an mRNA intermediate, as a portion of the poly A tail is retained and introns are not present. In this example, the direct repeats are highly conserved (identical in 12 of 13 bases), which, like the phylogenetic data, suggests that the insertion event took place in relatively recent evolutionary history. We conclude from this analysis that the insertion of the *Rpl37* retrogene into the 3′ UTR of a *Mus* ancestor, but not a shared ancestor with *Rattus*, created a novel 3′ splice acceptor site

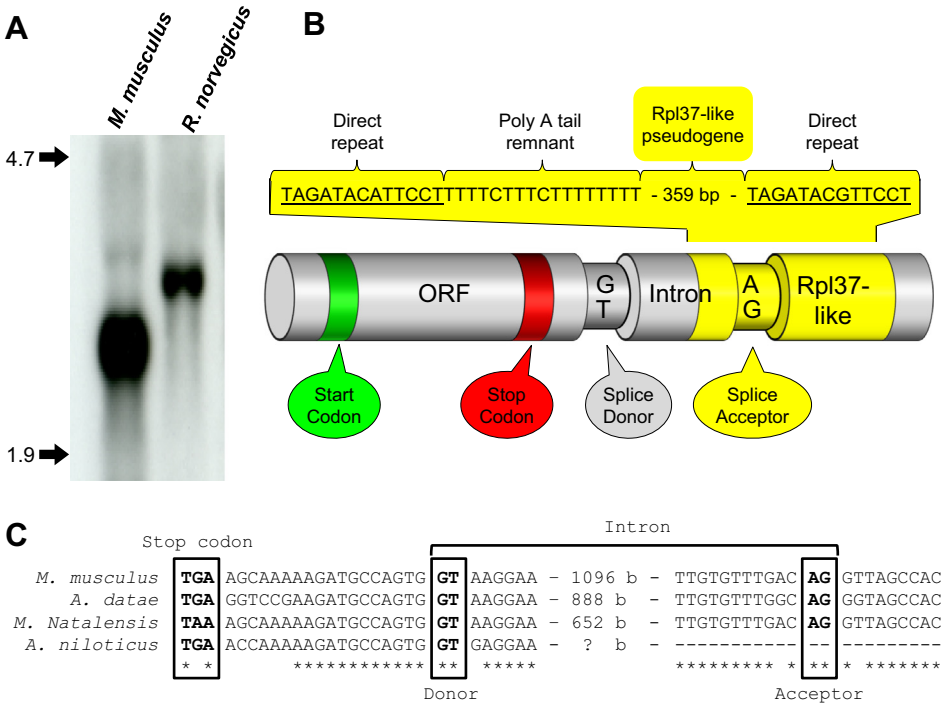

Fig. 2. Acquisition of *Rpl37*-like sequences and gain of an intron in *Mus musculus*. (A) Northern blot showing expression of Zfp36l3 mRNA in *Mus musculus* and *Rattus norvegicus* placentas. The blots were hybridized with random primed, ^{32}P dCTP-labeled DNA probes derived from the entire 2178 b of the *Mus musculus* open reading frame. Arrows indicate the positions of the 18S and 28S rRNAs at approximately 1.8 and 4.7 kb, respectively. (B) Diagram showing elements of the *Rpl37*-like sequence and the resulting intron in *Mus musculus Zfp36l3*. The presence of nearly identical flanking direct repeats, and remnants of a poly A tail, indicate that the *Rpl37*-like sequence was derived by retrotransposition. Note that the poly A tail remnant is shown here as poly T, because the sequence has been inserted into *Zfp36l3* in reverse orientation relative to the message sense. The splice donor site is located 18 base pairs after the *Zfp36l3* stop codon, and the splice acceptor site is provided by the *Rpl37*-like sequence. (C) Alignment of the potential splice donor and acceptor sites of the intron found in *Mus musculus Zfp36l3*, and the genes from the other species that have *Rpl37*-like sequences. We were unable to obtain the sequence of *Arvicanthis niloticus Zfp36l3* in the region containing the splice acceptor site, but based on its position in the phylogeny, it is predicted to possess the *Rpl37*-like sequences and presumably the intron.

that, together with an existing, endogenous, 5′ splice donor site, allowed for "intronization" of the 1096 b following the 5′ donor site (Fig. 2B).

These results suggest that the duplication event creating *Zfp36l3* in an ancestor of both *Mus musculus* and *Rattus norvegicus* occurred by the process of retrotransposition of the Zfp36l2 mRNA, and that a second, nested retrotransposition event occurred in an ancestor of *Mus musculus* that was not shared with *Rattus norvegicus*. We are unaware of any other examples of the Zfp36l2 mRNA serving as a retrotransposon in mammals (although see discussion of *N. gallili* above), but sequences related to the intronless Rpl37 mRNA occur frequently in the *Mus musculus* genome. In addition to its "true" genomic position, 115103182-115103521 of NC_000071.6 on mouse chromosome 5, a search of the *Mus musculus* reference genome on 8/17/14 identified many other sites with very high sequence similarities to Rpl37 mRNA, most of which appear to be pseudogenes.

To determine the phylogenetic distribution of this modification, we searched other genomic sequences from the family Muridae for the presence of Rpl37 mRNA sequences and splice sites. Although the presence of an intron at this site could not be verified due to a lack of placental RNA, *Zfp36l3* sequences from the Murinae species *Apomys datae* (Luzon montane forest mouse) and *Mastomys natalensis* (Natal multimammate mouse) also likely transcribe an mRNA with an intron in this location, since their genomic sequences also contain the *Rpl37* retrogene, and the splice donor and acceptor sites are conserved with *Mus musculus* (Fig. 2C). We were not able to extend the 3′ sequence of *Arvicanthis niloticus* (African grass mouse) genomic DNA far enough to determine whether it contains the *Rpl37* retrogene; however, based on its position in the phylogeny, we suspect that it will contain this sequence, and possibly the related intron.

3.4. Novel introns in Zfp36l3 in Cricetidae

A comparison of cDNA and genomic sequences from some Cricetidae species revealed another surprise: the presence of a new intron within the protein coding regions in some species. Sequencing of *Zfp36l3* from genomic DNA and placental cDNA from *Peromyscus maniculatus* (prairie deer mouse), a member of the North American subfamily Neotominae, revealed a 426 b intron in the C-terminal half of the protein coding region (Fig. S3A and B). Splicing occurs at standard GU/AG splice donor and acceptor sites. The spliced mRNA generates a predicted protein of 351 amino acids, less than half the length of the 725 amino acid *Mus musculus* ZFP36L3 protein. We sequenced the transcript, but not genomic DNA, from the closely related species *Peromyscus polionotus* (beach mouse), and splicing also takes place at this site in this species. *Zfp36l3* from *Reithrodontomys fulvescens* (fulvous harvest mouse), also a member of the Neotominae, contains conserved splice donor and acceptor sites, and there are in-frame stop codons within the predicted intron (Fig. S3A and B). We therefore predict that equivalent splicing occurs in *Reithrodontomys fulvescens*, resulting in a protein that is 330 amino acids in length. Similarly, *Nyctomys sumichrasti* (vesper rat), a member of the Central American Tylomyinae, contains the same splice donor and acceptor sites, and the putative intron contains a stop codon, making splicing in this species very likely (Fig. S3A and B). The protein sequences resulting from both the demonstrated and predicted splicing events just described were used in the alignment shown in Fig. S2.

These findings suggested that similar splicing of the Zfp36l3 transcript might also occur in another member of the Cricetidae, *Sigmodon hispidus* (cotton rat), based on its phylogenetic relation to the Neotominae and Tylomyinae (Fig. 1). However, sequencing of the open reading frame of a Zfp36l3 cDNA derived from placental RNA from *S. hispidus* revealed that the potential intron is not removed in this species. We confirmed this finding by northern blotting of placental RNA (data not shown). The Zfp36l3 transcript of *S. hispidus* encodes a large repeat region comprised of 15 tandem copies of a 19 amino acid sequence, positioned within the potential intron sequence. It therefore appears that, despite the presence of appropriate splice site sequences, this species has not spliced out an intron at this site, ultimately resulting in the expression of a protein of 730 amino acids (Fig. S3).

We also obtained placental cDNA sequences for Zfp36l3 mRNA from the Cricetinae species *Cricetulus griseus* (Chinese hamster) and *Mesocricetus auratus* (Syrian hamster). Splicing within the protein coding region of the Zfp36l3 transcript was not observed in either of these species (Fig. S3A and B), as determined by sequencing and northern blotting (data not shown). Although the same GU and AG splice donor and acceptor sites are present in all sequenced Cricetidae, only in Neotominae (the two *Peromyscus* species and *Reithrodontomys fulvescens*) is there a cytosine immediately 5′ of the splice acceptor, AG (Fig. S3A). This cytosine matches the preferred consensus for the major spliceosome (Burset et al., 2001), and may facilitate removal of the intron in these few species. However, the cytosine does not appear to be an absolute requirement, since splicing is also predicted in *Nyctomys sumichrasti*, which does not have this nucleotide. These data suggest that formation of this intron occurred approximately 17.5 million years ago (15.5–19.5, Fig. 1; Schenk et al., 2013) on the branch leading to the common ancestor of New World Cricetidae.

3.5. Evolutionary modifications in ZFP36L3

3.5.1. Repeat elements

There are several repeat elements in the ZFP36L3 proteins from the various species that are not found in the ancestral ZFP36L2 protein. In one such case, a short PQ-rich stretch, N-terminal of the TZF domain, has expanded in several of the species examined. In *Mus musculus* this sequence is QQKPKPQK, and this has increased in length, most notably in the Mongolian gerbil *Meriones unguiculatus*, the prairie vole *Microtus ochrogaster*, and Percival's spiny mouse *Acomys percivali*, by 7, 28, and 42 amino acids, respectively (Fig. S4). *Microtus ochrogaster* is a cricetid (Arvicolinae), while *Meriones unguiculatus* and *Acomys percivali* are murids (Gerbillinae and Deomyinae, respectively), suggesting that these expansions have taken place as independent events, an apparent example of parallel evolution. The mechanisms by which these expansions have formed are likely to be of the types used in the familiar triplet repeat diseases in man (Ashley and Warren, 1995; Liu and Wilson, 2012). Their contributions to the biochemical properties of the proteins are not known.

ZFP36L3 proteins from all members of Muridae, with the exception of *Meriones unguiculatus*, contain long, generally alanine-rich repetitive elements between the TZF domains and the C-termini. This sequence is highly variable in length, ranging from 208 amino acids in *Apomys datae* to 418 in *Mastomys natalensis* (Table S3), and therefore contributes to the major differences in protein lengths seen in the examined species. This alanine-rich repetitive element was first identified in mouse and rat (Blackshear et al., 2005; Frederick et al., 2008), and has been shown to be responsible for maintaining the *Mus musculus* protein in the cytosol (Frederick et al., 2008) (also see below). It is comprised of tandem repeats, usually of six to eight amino acids in length. Based on its distribution within the rodent phylogeny, the alanine-rich repetitive elements probably appeared in an ancestor of Muridae after it split from Cricetidae, approximately 20–23 million years ago.

These repeat regions differ in their physicochemical properties. For example, although the predicted isoelectric points of the repeat regions were generally basic, between 12 and 13, the values for *Arvicanthis niloticus* and *Acomys percivali* were 9.41 and 3 respectively (Table S3). This element was absent in most Cricetidae; however, ZFP36L3 from one member of this family, *Sigmodon hispidus*, contains a unique series of repetitive elements in essentially the same position as the alanine-rich repeat seen in Muridae. The length of the basic repeating unit of the *Sigmodon hispidus* element is 19 amino acids, which is much longer than the Muridae elements, and it has a more neutral pI of 6.95 (Table S3). The *Sigmodon hispidus* repetitive element shares no evident similarity in DNA or amino acid sequence to the alanine-rich element. These data suggest that the two types of repeats evolved independently in Muridae and Cricetidae. Possible mechanisms of the formation of tandem repeats and their importance in evolution have been reviewed recently (De Grassi and Ciccarelli, 2009; Gemayel et al., 2010; Hannan, 2012; Jankowski et al., 2000; Jansen et al., 2010).

3.5.2. Other sequence changes

In Fig. 3A, we have aligned the amino termini from the available ZFP36L3 sequences with ZFP36L2 sequences from the available rodent species, including several rodents that do not express ZFP36L3. The alignments of protein regions shown in Fig. 3A–C were extracted from a Clustal Omega alignment of the full-length ZFP36L3 proteins with the available full-length rodent ZFP36L2 proteins; the alignment of the full-length proteins is shown in Fig. S1. *Zfp36l3* sequences from all members of subfamily Murinae have gained downstream start codons, as compared to the other *Zfp36l3* and *Zfp36l2* sequences and, with the exception of *Batomys granti* (Luzon hairy-tailed rat), the original upstream start codons from this group have shifted out of frame. As a result, the Murinae species are predicted to encode ZFP36L3 proteins that are shorter at the N-terminus by approximately 21 amino acids, as compared to the other ZFP36L3 proteins and the parental ZFP36L2 proteins. *Batomys granti* is outside a clade ("core murines") that contains all of murine species sampled here (Fig. 1), and *Zfp36l3* from this species contains potential initiator methionines in both the longer and shorter positions.

Fig. 3B demonstrates the alignment of the TZF domains, and shows the great similarity among the ZFP36L2 and ZFP36L3 sequences at this key RNA binding domain. There are two sequence signatures characteristic of ZFP36L3 that are useful in distinguishing it from its family members, both of which were recognized in the initial descriptions of the mouse and rat protein (Blackshear et al., 2005; Frederick et al., 2008). One is found immediately before the second zinc finger; there is a proline (P) at this position in the ZFP36L3 proteins (with one exception, see below), while the other family members have a leucine (L) at this position (Fig. 3B). In most instances, the P residue of ZFP36L3 is encoded by CCT, while the corresponding L residue of ZFP36L2 is encoded by CTG. In an exception, ZFP36L3 from *Sigmodon hispidus* appears to have regained an L in this position by a single point mutation: CCT to CTT. This is an alternate codon for L, as compared to the one found in the ancestral ZFP36L2, and differs from the codon for proline by only one nucleotide. This suggests that an *S. hispidus* ancestor previously possessed the CCT codon, encoding proline at this position. In a second exception, *Acomys percivali* has apparently acquired a silent mutation; CCT to CCG, both of which encode proline. Collectively, these data suggest that the change to proline occurred in an early ancestor of the modern *Zfp36l3* genes. To determine whether this modification affected RNA binding, we previously made this L to P modification in human TTP, and expressed it in human embryonic kidney (HEK) 293 cells. This modification did not affect the

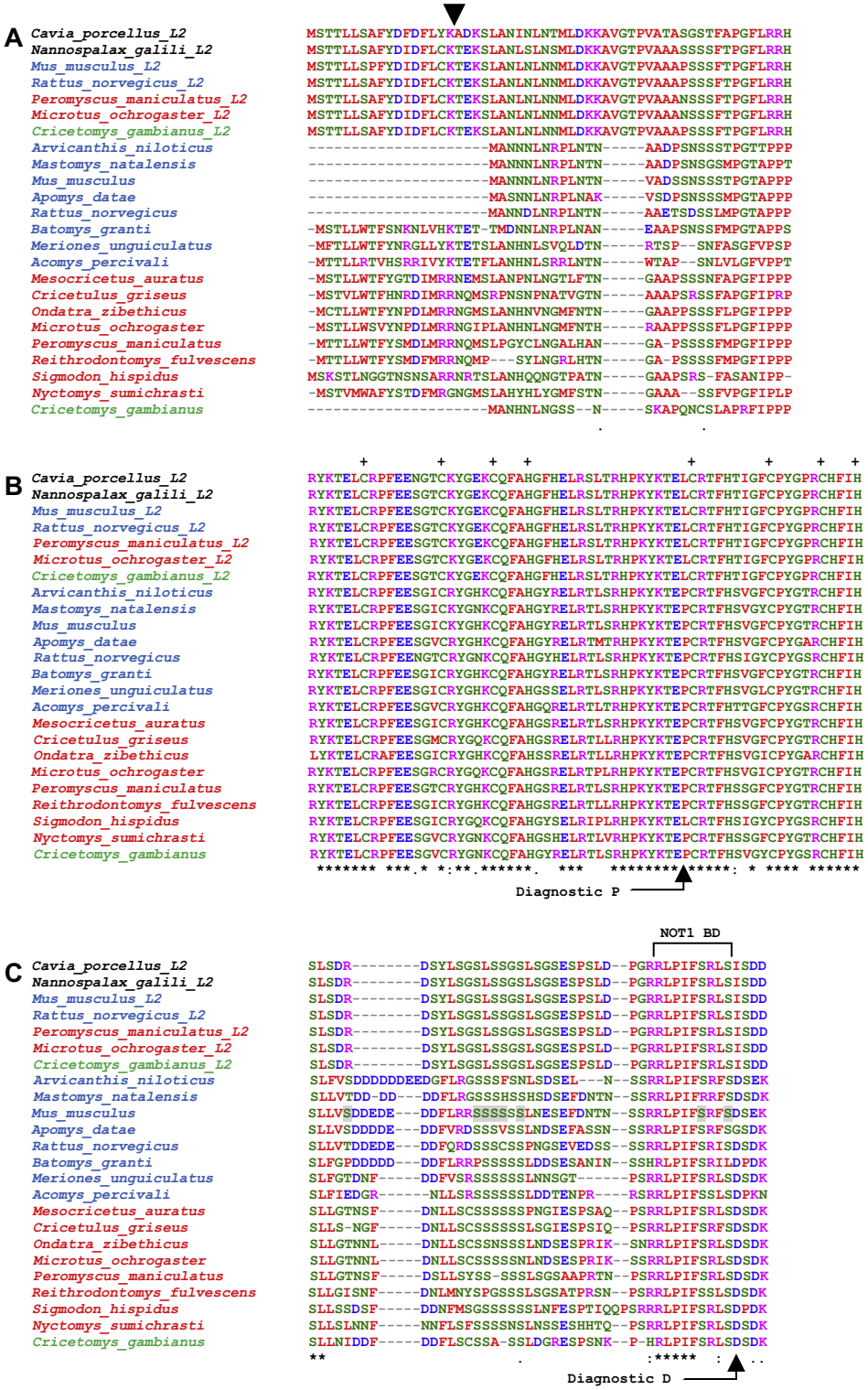

Fig. 3. Alignments of key protein domains in rodent ZFP36L2 and ZFP36L3. Shown are portions of several rodent ZFP36L2 proteins, and the analogous regions of the ZFP36L3 protein sequences discussed in this paper. We excluded ZFP36L3 sequences from *Peromyscus polionotus* and various members of the *Rattus* family, because of very high similarities to *Peromyscus maniculatus* and *Rattus norvegicus*, respectively. The original alignment was made using the entire proteins, with Clustal Omega and its default settings, and the three regions of the proteins shown are from the original alignment of the full-length proteins shown in Fig. S2. Species names in blue type are members of Muridae family; red are Cricetidae; the single green species is from Nesomyidae, and in black are two representative rodent species that do not have *Zfp36l3*. L2 refers to ZFP36L2; all other sequences are from ZFP36L3 from various species. The asterisks indicate amino acid identity at that site; the colons indicate amino acid similarity; and the periods indicate lesser similarity, according to the Clustal conventions. The location of the single intron in mouse and rat *Zfp36l2* is indicated by the arrowhead in A. The locations of the "diagnostic" P and D residues in ZFP36L3 are indicated in B and C, respectively, and the probable NOT1 binding site near the C-terminus in is indicated in C (NOT1 BD). In C, the shaded residues represent probable phosphorylation sites in *Mus musculus* ZFP36L3, except that, using the numbering system in NP_001009549.1, only two of the four serines highlighted at positions 694–697 are phosphorylated, and probably only one of the two residues at 718 and 721 is phosphorylated. See Section 3.5.2 for further details.

binding of human TTP to an RNA probe based on the TNF 3′ UTR ARE, either in maximum binding, or binding affinity (Lai et al., 2013a).

The second "diagnostic" modification is an aspartic acid residue immediately before the C-terminus of the protein, in place of the branched chain amino acid that is present in all other family members from mouse and rat (Fig. 3C). Aspartic acid at this position destroys the nuclear export sequence present in the *Mus musculus* ZFP36L1 and ZFP36L2 proteins (Frederick et al., 2008; Phillips et al., 2002). The aspartate residue is present in the ZFP36L3 proteins from all species shown in the alignment, except *Apomys datae*, in which there is a glycine at that position. This should still be incapable of supporting a functional nuclear export sequence. In one additional exception, ZFP36L3 from *Rattus tiomanicus* (Malayan field rat; not shown in the alignment) has a glutamic acid in the equivalent position. Like aspartate, this amino acid is acidic and should destroy the NES. These data suggest that all members of ZFP36L3 that have been sequenced to date are very unlikely to have a functioning nuclear export sequence at that site, in contrast to the presumed ancestral ZFP36L2 protein.

Also illustrated in Fig. 3C is another recently identified domain that almost completely overlaps with the non-functional NES. This domain is highly conserved among the mammalian gene family members, and is also found in related proteins from most eukaryotic lineages (Blackshear and Perera, 2014). This domain in human TTP was demonstrated to bind to the scaffolding protein NOT1 (Fabian et al., 2013). NOT1 is part of a complex that contains a number of proteins, including 3′–5′ exoribonucleases. Disrupting the interaction between NOT1 and TTP impairs TTP-mediated mRNA deadenylation (Fabian et al., 2013). The high degree of conservation of this domain suggests that similar interactions occur between NOT1 and the other family members, including all of the ZFP36L3 proteins sequenced to date and examined here (Blackshear and Perera, 2014).

Fig. 3C also shows the locations of phosphorylated serines previously identified in a mass spectrometry analysis of mouse ZFP36L3 purified after its expression in HEK 293 cells (Frederick, 2009). In many of the cases shown, the serines phosphorylated in the mouse are conserved among the various ZFP36L3 proteins, raising the possibility that many of the proteins from other species are phosphorylated at those sites. The effects of these phosphorylation events are not known, but in the mouse, they contribute to the appearance of the immunoreactive protein as two bands on SDS polyacrylamide gels, and the extreme C-terminal phosphorylation sites would be expected to modify the interaction with NOT1 proteins, as discussed previously by Fabian et al. (Fabian et al., 2013).

3.6. Subcellular localization of the ZFP36L3 proteins

The other three members of the TTP protein family in *Mus musculus* are nucleo-cytoplasmic shuttling proteins, with specific nuclear import and export sequences. In all cases, nuclear export can be blocked with leptomycin B (LMB) (Phillips et al., 2002). These characteristics extend to the single member of this protein family expressed in *Drosophila melanogaster*, Tis11, suggesting that they occurred in an ancient ancestral protein for the entire family (Twyffels et al., 2013). We showed earlier that mouse ZFP36L3, like the other mouse family members, possesses a nuclear localization signal between the tandem zinc fingers (Frederick et al., 2008). However, although the ZFP36L1 and ZFP36L2 proteins both contained highly conserved nuclear export sequences near their C-termini (Phillips et al., 2002), this sequence has been changed in mouse and rat ZFP36L3, as discussed in Section 3.5.2, so that this element is non-functional as a nuclear export sequence. Instead, the long alanine-rich repeat domain in mouse ZFP36L3 serves to maintain the protein in the cytosol, making the protein in this species permanently cytosolic (Frederick et al., 2008). This raises an important question: If Cricetidae lack the alanine-rich repeat region required to keep the protein in the cytosol in *Mus musculus*, how are they accomplishing this feat, presuming that a cytosolic location is necessary for the mRNA degrading functions of the protein?

We addressed this question with ZFP36L3 from *Peromyscus maniculatus*. In this species, the protein is relatively short, 350 amino acids, compared to 725 in *Mus musculus*, with the difference largely accounted for by the gain of the 265 amino acid alanine-rich repeat element in the *Mus musculus* protein, and the loss of intron sequences accounting for approximately 142 amino acids in the *Peromyscus maniculatus* protein. We generated mammalian expression plasmids with GFP fused to both the N and C-termini of *P. maniculatus* ZFP36L3, and transfected these into HEK 293 cells, which do not express significant levels of any TTP family members. We also raised polyclonal antisera to a recombinant fragment of ZFP36L3 from *P. maniculatus*. We found that both types of *P. maniculatus* ZFP36L3-GFP fusion proteins were localized to the cytosol, as was the untagged protein, localized by immunostaining (Fig. 4A and B). Somewhat surprisingly, both the tagged and untagged ZFP36L3 proteins were localized to the nucleus after LMB treatment, indicating that, in contrast to the mouse, *P. maniculatus* ZFP36L3 is a nucleocytoplasmic shuttling protein (Fig. 4A and B). The previously reported cytosolic localization of ZFP36L3 from *M. musculus*, and its failure to localize to the nucleus after LMB treatment (Frederick et al., 2008), were confirmed in this study (Fig. 4A).

The subcellular localization of the *P. maniculatus* ZFP36L3 protein was further studied by site directed mutagenesis. Using the nuclear export sequence prediction program NetNES 1.1 (la Cour et al., 2004), we identified two potential NESs in *P. maniculatus* ZFP36L3; one near amino acid 10, and one near amino acid 301. Several amino acids around each of these were deleted by site directed mutagenesis in the ZFP36L3-GFP expression plasmids, singly and together. When these plasmids were transfected into HEK 293 cells, deletion of the central seven amino acids of the potential C-terminal NES caused the fusion protein to localize to the nucleus, whereas deletion of the N-terminal NES had no effect (Fig. 4C). The C-terminal deletion removes two of the four hydrophobic residues that form the NES consensus proposed by la Cour et al. (2004). In most of the sequenced ZFP36L3 proteins, the four hydrophobic residues critical to the NES were conserved (Fig. 5), suggesting that the active NES in *P. maniculatus* might also be active in other species.

4. Discussion

We determined from these studies that *Zfp36l3* is present only in some families of the rodent superfamily Muroidea: Muridae, Cricetidae, and Nesomyidae. Based on recent phylogenetic estimates, this indicates that the gene first appeared in an ancestor common to these families between 23.7 and 37.8 million years ago. Our analyses suggest that this new gene arose as a retrogene, inserted into the X chromosome, that derived from the mRNA of the TTP family member *Zfp36l2*, which is expressed in all vertebrates examined to date. This conclusion is strongly supported by the sequence similarity between the ZFP36L3 and ZFP36L2 proteins, and the lack of the characteristic 5′ *Zfp36l2* intron in *Zfp36l3*. We could not precisely establish the boundaries of the retrotransposon insertion, although a rough approximation is clearly visible; the region of alignment between the *Zfp36l3* gene and the Zfp36l2 transcript starts at the 5′ end of the Zfp36l2 mRNA and continues into its 3′ UTR. After this presumably random insertion, the

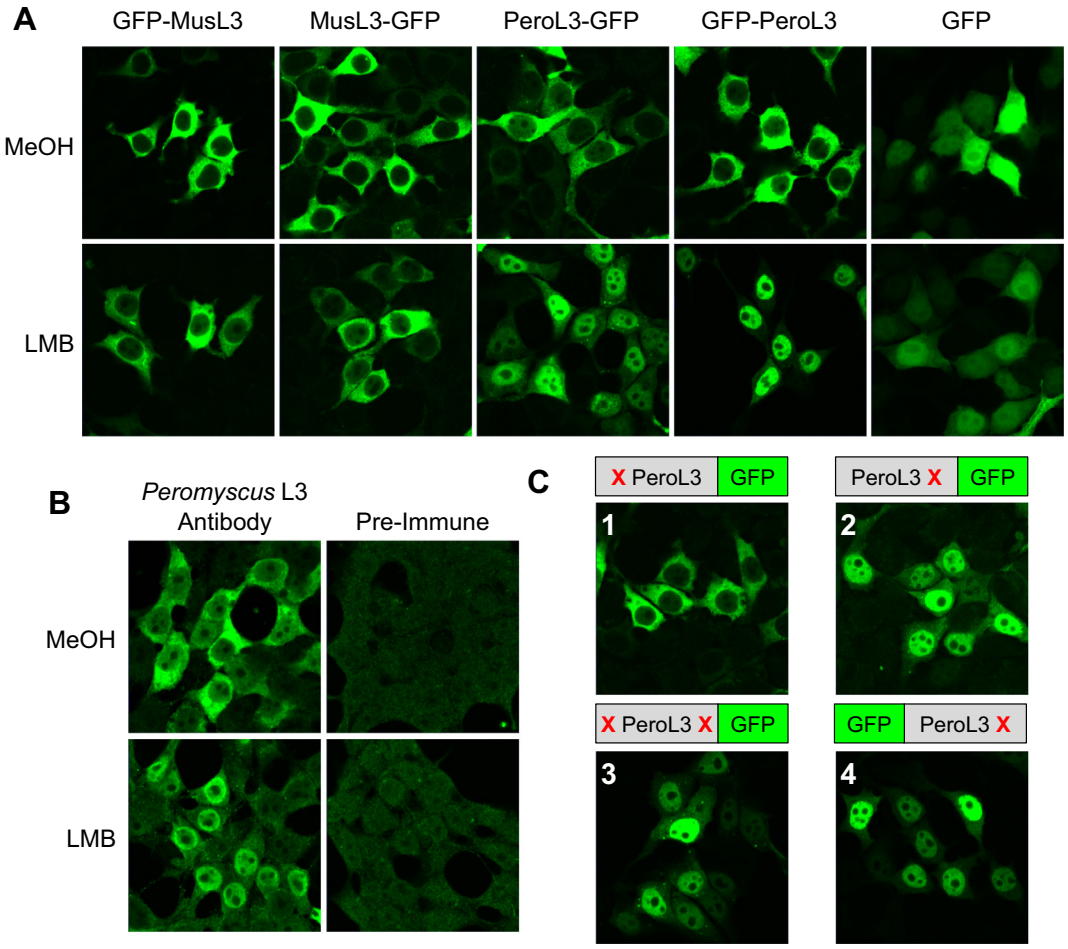

Fig. 4. Subcellular localization of ZFP36L3 from *Mus musculus* and *Peromyscus maniculatus*. Plasmids expressing ZFP36L3 proteins were transfected into HEK 293 cells and localized using GFP tags, or by immunostaining, with confocal microscopy. Where indicated, cells were treated with leptomycin B (LMB), an inhibitor of the nuclear export factor CRM1, or with vehicle (MeOH), for 3 h. (A) Cells were transfected with plasmids expressing *Mus musculus* or *Peromyscus maniculatus* ZFP36L3, with GFP fused to either the N- or C-terminal ends. *Mus musculus* ZFP36L3 and *Peromyscus maniculatus* ZFP36L3 are abbreviated to MusL3 and PeroL3, respectively. The end of ZFP36L3 to which GFP was fused is indicated by the order in which they are written. In the top row of (A) are shown cells treated with vehicle only, with GFP expression mainly in the cytosol in all cases, except in the case of GFP alone, which was expressed in both the nucleus and cytosol. In the bottom row of (A), after treatment with LMB, *Mus musculus* ZFP36L3 remained primarily cytosolic, regardless of which end was attached to GFP, whereas *Peromyscus maniculatus* ZFP36L3 was largely nuclear, indicating that the latter is a shuttling protein. Similar findings were observed with both N- and C-terminal tagging of both proteins. (B) Untagged *Peromyscus maniculatus* ZFP36L3 was transfected and localized by immunostaining, using an antibody (#2050) raised against a fragment of this protein. The *Peromyscus maniculatus* ZFP36L3 protein was again retained in the nucleus in response to LMB. The specificity of the antibody is shown using pre-immune serum under identical conditions (right panel in Fig. 4B). (C) Mutations that removed the N- and C-terminal predicted NESs were made to the *Peromyscus maniculatus* ZFP36L3 – GFP fusion constructs. The diagram above each image indicates the end to which GFP was fused and which predicted NES was deleted. Images 2 and 4 indicate that deletion of the C-terminal NES alone results in accumulation of the protein in the nucleus with both of the GFP tagged proteins, while image 1 shows that deletion of the predicted N-terminal NES has no effect on localization. Proteins containing both putative NES deletions were also retained in the nucleus (image 3).

ancestral rodent would have acquired promoter sequences through a variety of potential mechanisms. The presence of *Zfp36l3* ever since in Muridae, Cricetidae and Nesomyidae strongly suggests that this gene has provided a useful function for the descendant species. The nature of this advantage is suggested by the phenotype of the *Zfp36l3* KO mice, which exhibit decreased neonatal survival and altered placental gene expression (D.J. Stumpo and P.J. Blackshear, manuscript submitted).

In humans and chimpanzees, the rate of novel gene retrocopy insertions has been estimated at one event for every 6000 individuals (Ewing et al., 2013). The most common fate of newly formed retrogenes is pseudogenization (Kaessmann, 2010; Lynch, 2002; Lynch and Conery, 2000). In general, a duplicated paralog, identical to the parental gene, serves no purpose and will eventually be lost by entropic decay; therefore, random acquisition of advantageous mutations must occur before the gene is lost (Force et al., 1999; Nowak et al., 1997; Sidow, 1996). Such events have presumably occurred in the case of *Zfp36l3* to allow its continued existence in the Muridae, Cricetidae and Nesomyidae, in addition to the fortuitous acquisition of promoter elements. Two noteworthy mutations that appear to have taken place in one of the earliest ancestral ZFP36L3 proteins include: conversion of a leucine residue to a proline in the lead in sequence of the second zinc finger, and functional loss of the C-terminal NES. Loss of this NES, which was possibly the only NES in the ancestral ZFP36L3, based on its origin as ZFP36L2, would have resulted in localization of the protein to the nucleus, due to the presence of a nuclear localization sequence in the inter-finger region. To maintain its cytoplasmic functionality, ZFP36L3 would have had to acquire either a new NES, which it appears to have done in several species, or a new element that by itself was capable of maintaining cytoplasmic localization, as has occurred in many of the Muridae. Both modifications appear to have occurred as separate evolutionary events after what we presume to be the initial loss of the C-terminal NES. Changes in subcellular localization of paralogs have been studied in yeast, where it was observed that 24–37%

Fig. 5. Core residues of the novel *Peromyscus maniculatus* ZFP36L3 NES are conserved in other ZFP36L3 proteins, but not in ZFP36L2. Shown are sections of the ZFP36L2 and ZFP36L3 protein alignment from Fig. S2 that correspond to the newly identified NES from *P. maniculatus*. The alignment was performed with Clustal Omega, and is formatted according to Clustal conventions, as described in the legend to Fig. 3. A general consensus for a leucine-rich NES has been proposed by la Cour et al. (2004): [LIVFM]$X_{(2\ or\ 3)}$[LIVFM]$X_{(2\ or\ 3)}$[LIVFM]X[LIVFM], where letters within the square brackets are critical hydrophobic residues, any one of which may be present at that position, X indicates any amino acid, and subscript values indicate the number of amino acids. The NES that we identified in *P. maniculatus* follows the pattern $LX_3LX_2LX_1G$, which matches the consensus (gray highlight), with the exception of the last residue, which is a G (cyan highlight). This suggests that the G residues found in this position, in several of the other species (cyan highlight), might also support a functional NES. Alternatively, the immediately preceding residue in all species is either an F or L, both hydrophobic amino acids known to contribute to NESs at other sites. Virtually all of the remaining core residues in the ZFP36L3 proteins from the other species match the NES consensus (gray highlight). In the only exceptions, there was an S, or an N present at the fourth position. We did not examine what effect this might have on the activity of those putative NESs. None of the ZFP36L2 proteins contain sequences resembling the NES consensus at this position, although the sequence DSL, found in the middle of the known *P. maniculatus* NES, is completely conserved in all ZFP36L2 and ZFP36L3 proteins sequenced to date. The position of this short sequence is indicated by the asterisks below the alignment. Underlined residues in the *P. maniculatus* sequence indicate the seven amino acids that were deleted in ZFP36L3-GFP fusion constructs that destroyed the function of the novel NES. Species names in blue type are members of Muridae family; red are Cricetidae; the single green species is in Nesomyidae; and in black are two representative rodent species that do not have *Zfp36l3*.

of duplicate gene products localize, partly or completely, to distinct subcellular areas (Marques et al., 2008).

The appearance of a new expressed protein coding gene by retrotransposon insertion is a relatively uncommon event. In some cases, the same retrogene may insert in many places throughout the genome. Examination of the genomes of *Mus musculus* and *Rattus norvegicus* suggests that the *Zfp36l2* retrogene insertion that gave rise to *Zfp36l3* has occurred only once. In contrast, *Rpl37* is an example of a gene for which many retropseudogene copies exist throughout the *Mus musculus* genome. One pseudogenized copy of this latter gene has inserted into the 3′ UTR of *Zfp36l3* in an ancestor to a subgroup of Murinae. The *Zfp36l3* locus is therefore comprised of nested retrotransposons in the species descending from this ancestor. Interestingly, the *Rpl37* retrogene insertion in *Mus musculus* provides a novel splice acceptor site, which, together with an endogenous splice donor site, results in the formation of an intron in the 3′ UTR of *Zfp36l3*. The *Rpl37* retrogene and presumably the intron it is responsible for are also present in the "core" murine species (Rowe et al., 2008) *Apomys datae*, *Mastomys natalensis* and probably *Arvicanthis niloticus*, but are not present in the more distantly related *Rattus norvegicus*. This dates the first appearance of this element to a very short branch at approximately 12 million years ago (Schenk et al., 2013). The lack of an intron in *R. norvegicus* explains its much greater mRNA length, as confirmed by direct sequencing of cDNA and northern blotting analysis. Introns situated more than 50–55 bases past the stop codon are expected to trigger nonsense mediated decay (NMD), a mechanism that degrades transcripts containing premature stop codons (Lewis et al., 2003; Nagy and Maquat, 1998). However, the intron found in the 3′ UTR of *Mus musculus* and related species begins only 18 b past the stop codon and therefore would not be expected to activate the NMD pathway.

We also identified a second intron in *Zfp36l3*, within the protein coding region, in some members of Cricetidae. Examples of this type of *de novo* intron "invention" are extremely rare, with few known occurrences in placental mammals (Kordis, 2011; Szczesniak et al., 2011). Our data and analysis indicate that splicing of this intron occurs in the New World subfamilies Neotominae and Tylomyinae. This localizes the evolutionary origin of the coding region intron to a branch less than 500,000 years long at about 12.8 million years ago (Schenk et al., 2013). We expected to see splicing of this intron in *Sigmodon hispidus* of the Sigmodontinae, based on its relation to Neotominae and Tylomyinae. However, sequencing and northern blots revealed that splicing does not occur in this species; instead, the *S. hispidus Zfp36l3* gene encodes a repeat region composed of 15 tandem copies of a 19 amino acid sequence. The function of this repetitive sequence is unknown, but it may play a role in cytoplasmic localization of the protein, as in *M. musculus*. In the species that do contain this intron, the overall size of the predicted ZFP36L3 proteins is decreased, with the proteins from *Reithrodontomys fulvescens* and *Peromyscus maniculatus* being only 330 and 350 amino acids long, respectively, whereas the proteins from *Arvicanthis niloticus* and *Mastomys natalensis*, which do not contain coding region introns and which have long alanine-rich repetitive elements, are 790 and 882 amino acids long, respectively.

The protein sequences encoded by *Zfp36l3* were extremely variable in the species we examined, particularly compared to the almost complete lack of sequence divergence in the parental ZFP36L2 proteins. This variability included differences in the locations of the initiator methionines, and the presence and number of repetitive elements. Another key difference between ZFP36L2 and ZFP36L3 from *Mus musculus* and *Rattus norvegicus* was found in the highly conserved lead-in sequence to the second zinc finger: KYKTEL in ZFP36L2 was KYKTEP in ZFP36L3 in both species. We used this change as a "diagnostic" indicator of ZFP36L3 because a leucine is present at this position in all mammalian family members other than ZFP36L3 that we have examined to date. Although we have not observed a proline residue in this position in any mammalian family members other than ZFP36L3, a proline at this site is found in some family member proteins from fungi (see, for example, Wells et al., 2015). Direct examination has shown that introducing the L to P modification into human TTP does not change its ability to interact with RNA (Lai et al., 2014). The proline residue was present in this position in ZFP36L3 proteins in all species examined, with the exception of *Sigmodon hispidus*, in which the original leucine has been restored, presumably by a mutation changing the CCT proline codon found in the other species to CTT.

The second sequence characteristic of ZFP36L3 proteins from *M. musculus* and *R. norvegicus* that we have used as a diagnostic marker is a change from isoleucine to aspartate in a C-terminal sequence that is highly conserved among metazoan TTP family proteins (Blackshear and Perera, 2014). As noted in 3.52, this change disrupts the ability of this conserved sequence to serve as

a leucine-rich nuclear export sequence (Frederick et al., 2008). The characteristic aspartate residue was found in all sequences examined here, with two exceptions: There were glutamic acid and glycine residues at that site in *Rattus tiomanicus* and *Apomys datae*, respectively. However, these changes would not be expected to restore the function of the nuclear export sequence.

One of the most striking examples of protein sequence variability in the ZFP36L3 proteins examined here were the differences in the presence and length of the alanine-rich repeat region. We previously showed, in *M. musculus* ZFP36L3, that the long alanine-rich repeat region was necessary to maintain localization of the protein in the cytoplasm (Frederick et al., 2008). These results were confirmed in the current study. Similar alanine-rich repeat regions were found in other Muridae species; these varied greatly in length and composition, and contributed substantially to the large differences in protein size observed among the various species. Repeating sequences have the capacity to change rapidly in length and sequence composition, and it appears likely that the DNA encoding the repeat elements present in *Zfp36l3* originated *de novo* through expansion of a shorter sequence. Misalignment of DNA strands during meiosis, or during replication (Gemayel et al., 2010; Jankowski et al., 2000; Jansen et al., 2010), are potential mechanisms for the expansion and contraction of repeat sequences. Blocks of identical or nearly identical DNA are clearly evident in the region comprising the repeat elements, and are consistent with a misalignment-based mechanism of expansion.

A striking finding from the analysis of the protein sequences from the Cricetidae is that many of these proteins entirely lacked similar alanine repeat sequences, resulting in considerably smaller overall protein size. The absence of alanine-rich repeats, coupled with the lack of a functioning C-terminal NES, suggested the possibility that these proteins might be localized to the nucleus. This would result in a complete change in role, since the TTP family members, including ZFP36L3, function to promote the decay of AU-rich sequence-containing mRNAs in the cytoplasm. However, we demonstrated by transfection experiments that ZFP36L3 from *Peromyscus maniculatus*, either native, or GFP-tagged on either end, was almost entirely expressed in the cytoplasm, but could be trapped in the nucleus using the nuclear export blocker leptomycin B. These results identified ZFP36L3 as a nucleocytoplasmic shuttling protein in this species, in contrast to its full-time cytoplasmic status in *M. musculus*. As described in Section 3.6, this led to the discovery of a novel, functional nuclear export sequence in this protein, which was present in many of the related cricetid species.

Many questions concerning *Zfp36l3* remain unanswered. For example, we do not know the mechanism of the extreme tissue specificity of expression, at least in *Mus musculus*, where the expression is limited to the placenta and yolk sac. Comparative promoter and epigenetic analyses may be useful in this regard. A second critical question is whether the presumed functions of the ZFP36L3 protein in placental and yolk sac physiology are compensated for by one or more of the other members of the TTP family in species that do not express ZFP36L3, which include many rodents as well as all non-rodent mammals, including man.

5. Conclusion

Zfp36l3 is a member of a small family of RNA binding proteins that is found exclusively in rodents of the Muroidea superfamily. The gene appears to have arisen by retrotransposition of an mRNA encoded by a related gene, *Zfp36l2*, into the ancestral X chromosome between 23.7 and 37.8 million years ago. The orthologues of this gene described here illustrate a variety of mechanisms by which new genes evolve. We identified examples of new intron formation through both *de novo* invention of splice sites and through a retrotransposon-supplied splice acceptor site. Within the predicted proteins, we found several instances of emergence of repetitive elements, some of which had a bearing on the overall size of the proteins and/or influenced subcellular localization. The *de novo* formation of a new nuclear export sequence, found in one species, and predicted in several others, was also characterized. There were also differences in start codon utilization, causing the ZFP36L3 proteins to be shorter at their N-terminal ends in members of the Murinae clade. Outside of the highly conserved tandem zinc finger domain and the putative C-terminal NOT1 binding domain, the various proteins have undergone marked changes throughout their lengths. *Zfp36l3* is a recently acquired, rapidly evolving gene, and the various orthologues described here contain sequences that illustrate various mechanisms by which new genes emerge and evolve.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health, NIEHS, and the National Science Foundation grant DEB-0841447 to S.J.S. We are grateful to the many colleagues who provided DNA or tissue samples, as listed in Table S4. We thank Lori Edwards and Dr. Robert Petrovich for the peptide expression and purification, Dr. Brian Bennett for help with the *Nannospalax galili* sequences, Dr. Agnes Janoshazi and Jeff Tucker for help with the confocal microscopy, and Drs. David Miller and Ron Cannon for helpful comments on the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.10.016.

References

Ashley Jr., C.T., Warren, S.T., 1995. Trinucleotide repeat expansion and human disease. Annu. Rev. Genet. 29, 703–728.
Blackshear, P.J., 2002. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. Biochem. Soc. Trans. 30, 945–952.
Blackshear, P.J., Perera, L., 2014. Phylogenetic distribution and evolution of the linked RNA-binding and NOT1-binding domains in the tristetraprolin family of tandem CCCH zinc finger proteins. J. Interferon Cytokine Res. 34, 297–306.
Blackshear, P.J., Phillips, R.S., Ghosh, S., Ramos, S.B.V., Ramos, S.V.B., Richfield, E.K., Lai, W.S., 2005. Zfp36l3, a rodent X chromosome gene encoding a placenta-specific member of the Tristetraprolin family of CCCH tandem zinc finger proteins. Biol. Reprod. 73, 297–307.
Brooks, S.A., Blackshear, P.J., 2013. Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. Biochim. Biophys. Acta 1829, 666–679.
Burset, M., Seledtsov, I.A., Solovyev, V.V., 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res. 29, 255–259.
Carrick, D.M., Lai, W.S., Blackshear, P.J., 2004. The tandem CCCH zinc finger protein tristetraprolin and its relevance to cytokine mRNA turnover and arthritis. Arthritis Res. Ther. 6, 248–264.
De Grassi, A., Ciccarelli, F.D., 2009. Tandem repeats modify the structure of human genes hosted in segmental duplications. Genome Biol. 10, R137.
Ewing, A.D., Ballinger, T.J., Earl, D., Harris, C.C., Ding, L., Wilson, R.K., Haussler, D., 2013. Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. Genome Biol. 14, R22.
Fabian, M.R., Frank, F., Rouya, C., Siddiqui, N., Lai, W.S., Karetnikov, A., Blackshear, P. J., Nagar, B., Sonenberg, N., 2013. Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. Nat. Struct. Mol. Biol. 20, 735–739.
Fabre, P.-H., Hautier, L., Dimitrov, D., Douzery, E.J.P., 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. BMC Evol. Biol. 12, 88.
Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.
Frederick, E.D., 2009. ZFP36L3: A Unique Member of the Tristetraprolin Family of RNA-Binding CCCH Tandem Zinc Finger Proteins. Dissertation. Duke University, Durham.

Frederick, E.D., Ramos, S.B.V., Blackshear, P.J., 2008. A unique C-terminal repeat domain maintains the cytosolic localization of the placenta-specific tristetraprolin family member ZFP36L3. J. Biol. Chem. 283, 14792–14800.
Gemayel, R., Vinces, M.D., Legendre, M., Verstrepen, K.J., 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. 44, 445–477.
Hannan, A.J., 2012. Tandem repeat polymorphisms: mediators of genetic plasticity, modulators of biological diversity and dynamic sources of disease susceptibility. Adv. Exp. Med. Biol. 769, 1–9.
Huchon, D., Chevret, P., Jordan, U., Kilpatrick, C.W., Ranwez, V., Jenkins, P.D., Brosius, J., Schmitz, J., 2007. Multiple molecular evidences for a living mammalian fossil. Proc. Natl. Acad. Sci. U.S.A. 104, 7495–7499.
Jankowski, C., Nasar, F., Nag, D.K., 2000. Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. Proc. Natl. Acad. Sci. U.S.A. 97, 2134–2139.
Jansen, A., Gemayel, R., Verstrepen, K.J., 2010. Unstable microsatellite repeats facilitate rapid evolution of coding and regulatory sequences. Genome Dyn. 7, 108–125.
Kaessmann, H., 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20, 1313–1326.
Kaessmann, H., Vinckenbosch, N., Long, M., 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 10, 19–31.
Kordis, D., 2011. Extensive intron gain in the ancestor of placental mammals. Biol. Dir. 6, 59.
la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K., Brunak, S., 2004. Analysis and prediction of leucine-rich nuclear export signals. Protein Eng. Des. Sel. 17, 527–536.
Lai, W.S., Carballo, E., Strum, J.R., Kennington, E.A., Phillips, R.S., Blackshear, P.J., 1999. Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. Mol. Cell. Biol. 19, 4311–4323.
Lai, W.S., Carballo, E., Thorn, J.M., Kennington, E.A., Blackshear, P.J., 2000. Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. J. Biol. Chem. 275, 17827–17837.
Lai, W.S., Perera, L., Hicks, S.N., Blackshear, P.J., 2014. Mutational and structural analysis of the tandem zinc finger domain of tristetraprolin. J. Biol. Chem. 289, 565–580.
Lai, W.S., Perera, L., Hicks, S.N., Blackshear, P.J., 2013a. Mutational and structural analysis of the tandem zinc finger domain of tristetraprolin. J. Biol. Chem. 289, 565–580.
Lai, W.S., Stumpo, D.J., Kennington, E.A., Burkholder, A.B., Ward, J.M., Fargo, D.L., Blackshear, P.J., 2013b. Life without TTP: apparent absence of an important anti-inflammatory protein in birds. Am. J. Physiol. Regul. Integr. Comp. Physiol. 305, R689–R700.
Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R., Berns, A., 1991. Simplified mammalian DNA isolation procedure. Nucleic Acids Res. 19, 4293.
Lewis, B.P., Green, R.E., Brenner, S.E., 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc. Natl. Acad. Sci. U.S.A. 100, 189–192.
Liu, Y., Wilson, S.H., 2012. DNA base excision repair: a mechanism of trinucleotide repeat expansion. Trends Biochem. Sci. 37, 162–172.
Lynch, M., 2002. Genomics, gene duplication and evolution. Science 297, 945–947.
Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.
Marques, A.C., Vinckenbosch, N., Brawand, D., Kaessmann, H., 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol. 9, R54.
Nagy, E., Maquat, L.E., 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. 23, 198–199.
Nowak, M.A., Boerlijst, M.C., Cooke, J., Smith, J.M., 1997. Evolution of genetic redundancy. Nature 388, 167–171.
Phillips, R.S., Ramos, S.B.V., Blackshear, P.J., 2002. Members of the tristetraprolin family of tandem CCCH zinc finger proteins exhibit CRM1-dependent nucleocytoplasmic shuttling. J. Biol. Chem. 277, 11606–11613.
Rowe, K.C., Reno, M.L., Richmond, D.M., Adkins, R.M., Steppan, S.J., 2008. Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): multilocus systematics of the old endemic rodents (Muroidea: Murinae). Mol. Phylogenet. Evol. 47, 84–101.
Schenk, J.J., Rowe, K.C., Steppan, S.J., 2013. Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. Syst. Biol. 62, 837–864.
Sidow, A., 1996. Gen(om)e duplications in the evolution of early vertebrates. Curr. Opin. Genet. Dev. 6, 715–722.
Stumpo, D.J., Broxmeyer, H.E., Ward, T., Cooper, S., Hangoc, G., Chung, Y.J., Shelley, W.C., Richfield, E.K., Ray, M.K., Yoder, M.C., Aplan, P.D., Blackshear, P.J., 2009. Targeted disruption of Zfp36l2, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis. Blood 114, 2401–2410.
Stumpo, D.J., Byrd, N.A., Phillips, R.S., Ghosh, S., Maronpot, R.R., Castranio, T., Meyers, E.N., Mishina, Y., Blackshear, P.J., 2004. Chorioallantoic fusion defects and embryonic lethality resulting from disruption of Zfp36L1, a gene encoding a CCCH tandem zinc finger protein of the Tristetraprolin family. Mol. Cell. Biol. 24, 6445–6455.
Szczesniak, M.W., Ciomborowska, J., Nowak, W., Rogozin, I.B., Makalowska, I., 2011. Primate and rodent specific intron gains and the origin of retrogenes with splice variants. Mol. Biol. Evol. 28, 33–37.
Taylor, G.A., Carballo, E., Lee, D.M., Lai, W.S., Thompson, M.J., Patel, D.D., Schenkman, D.I., Gilkeson, G.S., Broxmeyer, H.E., Haynes, B.F., Blackshear, P.J., 1996. A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. Immunity 4, 445–454.
Twyffels, L., Wauquier, C., Soin, R., Decaestecker, C., Gueydan, C., Kruys, V., 2013. A masked PY-NLS in Drosophila TIS11 and its mammalian homolog tristetraprolin. PLoS ONE 8, e71686.
Vanin, E.F., 1985. Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. 19, 253–272.
Wells, M.L., Hicks, S.N., Perera, L., Blackshear, P.J., 2015. Functional equivalence of an evolutionarily conserved RNA binding module. J. Biol. Chem. 290, 24413–24423.