Published online 13 March 2023

Nucleic Acids Research, 2023, Vol. 51, No. 7 ***3185–3204***
https://doi.org/10.1093/nar/gkad152

PML and PML-like exonucleases restrict retrotransposons in jawed vertebrates

Sabateeshan Mathavarajah[1], Kathleen L. Vergunst[2], Elias B. Habib[1], Shelby K. Williams[2], Raymond He[2], Maria Maliougina[2], Mika Park[1], Jayme Salsman[1], Stéphane Roy[3], Ingo Braasch[4], Andrew J. Roger[2], David N. Langelaan[2,*] and Graham Dellaire[1,2,*]

[1]Department of Pathology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada, [2]Department of Biochemistry & Molecular Biology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada, [3]Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montréal, QB, Canada and [4]Michigan State University, Department of Integrative Biology and Ecology, Evolution, and Behavior Program, East Lansing, MI, USA

Received August 05, 2022; Revised February 16, 2023; Editorial Decision February 18, 2023; Accepted February 21, 2023

ABSTRACT

We have uncovered a role for the promyelocytic leukemia (*PML*) gene and novel PML-like DEDDh exonucleases in the maintenance of genome stability through the restriction of LINE-1 (L1) retrotransposition in jawed vertebrates. Although the mammalian PML protein forms nuclear bodies, we found that the spotted gar PML ortholog and related proteins in fish function as cytoplasmic DEDDh exonucleases. In contrast, PML proteins from amniote species localized both to the cytoplasm and formed nuclear bodies. We also identified the PML-like exon 9 (*Plex9*) genes in teleost fishes that encode exonucleases. Plex9 proteins resemble TREX1 but are unique from the TREX family and share homology to gar PML. We also characterized the molecular evolution of TREX1 and the first non-mammalian TREX1 homologs in axolotl. In an example of convergent evolution and akin to TREX1, gar PML and zebrafish Plex9 proteins suppressed L1 retrotransposition and could complement *TREX1* knockout in mammalian cells. Following export to the cytoplasm, the human PML-I isoform also restricted L1 through its conserved C-terminus by enhancing ORF1p degradation through the ubiquitin-proteasome system. Thus, PML first emerged as a cytoplasmic suppressor of retroelements, and this function is retained in amniotes despite its new role in the assembly of nuclear bodies.

INTRODUCTION

The *promyelocytic leukemia* (*PML*) gene was initially discovered as the oncogenic fusion partner of RARα in a translocation event that causes acute promyelocytic leukemia (1,2). In mammals, the PML protein forms a subnuclear organelle known as the PML nuclear body (NB) (1–3) that plays a prominent role in antiviral immunity and is a hub for the association and post-translational modification of over 150 proteins involved in transcriptional regulation, replication, and repair of DNA (3–5). PML is believed to contribute to tumour suppression through several mechanisms including the modulation of innate immune pathways as well as the regulation of the PTEN-AKT-mTOR signaling axis (3,6–9). Maintenance of genomic stability is also implicated in tumour suppression (10,11), and PML loss is associated with increased DNA damage and genome instability (5,8,12). Activation of the DNA damage response (DDR) triggers PML NB fission and the association of DNA repair factors with bodies, and loss or over-expression of PML inhibits DNA repair by homologous recombination (5,13,14). PML NBs also help maintain telomeric chromatin integrity in embryonic stem cells (15). However, a direct role for the PML protein itself in genome stability has remained elusive.

Endogenous retroelements such as the long-interspersed element-1 (LINE1/L1), are major contributors to genome instability and evolution (16,17). L1 elements are active autonomous retrotransposons, and almost one fifth of the human genome (17%) is comprised of L1 retroelement DNA (18). On average, human genomes contain 80–100 copies of retrotransposition-competent L1s, and these elements are responsible for the majority of our retrotransposition events (19–21). While L1s contribute a source of genetic variation, this is juxtaposed to how they corrupt genome integrity (17,22–24). For this reason, the human genome encodes numerous host factors that have been described as potent enhancers and suppressors of L1 activity (25,26). In addition to gene disruption by their insertion, L1 activation manifests in human disease through overstimulation of the

*To whom correspondence should be addressed. Tel: +1 902 494 8928; Email: dlangela@dal.ca
Correspondence may also be addressed to Dr. Graham Dellaire. Tel: +1 902 494 4730; Email: dellaire@dal.ca

DNA-sensing innate immune signalling pathways such as the cGAS-STING axis (27–29). Recent evidence has linked L1s as a driving force in autoimmune disease (29–34). For example, multiple genes associated with Aicardi-Goutières syndrome, an autoimmune disease sharing features with systemic lupus erythematosus, have been shown to encode regulators of L1 activity including *SAMHD1, ADAR1* and *RNASEH2* (32,33,35–37). Aicardi-Goutières syndrome is also caused by loss of DEDDh family *Three-prime repair exonuclease 1* (*TREX1*), a potent suppressor of L1 activity and an essential brake in the type I interferon responses that when impaired contributes to systemic autoimmunity (38). Although TREX1 is the predominant DNA exonuclease in mammals, non-mammalian TREX1 homologs have not been characterized in other vertebrates, even in species whose genomes encode both L1 and the related L2 retrotransposons such as the teleost fishes (39).

Teleost bony fish species genomes including zebrafish lack a *PML* ortholog, presumably because of the tetraploidization of their genomes and a multitude of gene rearrangement events that occurred sometime between 300 and 450 million years ago (40,41). In mammals, the functional diversity of the human PML protein is derived from its seven isoforms (PML-I to VII) that diverge at the C-terminus and are linked to specific cellular functions and protein interactions (14). However, most species of vertebrates including rodents, encode only a few PML isoforms, of which the most highly expressed resemble orthologs of the longest human isoform, PML-I. The human PML-I isoform is distinguished from other isoforms by encoding both a nuclear export signal (NES) and an uncharacterized C-terminal domain encoded by exon 9 that shares homology to the DEDDh family of exonucleases and in particular, TREX1 (42,43). Here, we use a molecular evolution approach to better understand the conserved function of PML-I in vertebrate genome stability, uncovering the convergent evolution of novel PML-like DEDDh exonucleases in different fish lineages and amniotes, as well as a primordial role for PML in the suppression of L1 retrotransposition, a known driver of autoimmune disease and genome instability in cancer (44,45).

MATERIALS AND METHODS

Phylogenetic analyses

Two rounds of searches on the protein and transcriptome sequences were carried out to detect homologs to the human PML sequence in sequenced eukaryote species. A phylogenetic dataset of PML proteins was constructed by collecting 72 protein sequences from NCBI via a PSI-BLASTp or tBLASTn search followed by reverse BLASTP. For searches, in the first round we targeted both the full length PML (sequence for isoform PML-I) and the C-terminus of PML (an amino acid stretch from positions 600–882; the CDE or C-terminal DEDDh exonuclease domain) in the second round. The logic behind this stems from the RBCC motif being highly conserved and associated with TRIM19 proteins. The C-terminus is unique to PML. The *E-value* threshold (-e) was set as 0.001, the *E*-value threshold for inclusion in the multipass model (-h) as 0.002 (default value), and the maximum running iterations (-j) as 5. For transcriptome sequences, tBLASTn was used and the *E-value* threshold (-e) was set as 0.001. For species with fully sequenced genomes available, PML orthology was confirmed by synteny to the human *PML* loci.

A separate dataset of the Plex and TREX exonucleases was constructed through a similar approach but searching for both the C-terminus of PML, zebrafish Plex9.1 and Plex9.2 and TREX1. resulting in a dataset of 104 proteins. Previously identified homolog PML-CDE sequences that had homologous sequence and synteny were added to the Plex and TREX dataset. Both datasets were aligned using mafft einsi with default settings (46). The ExoIII dataset was trimmed so that only the ExoIII, and ExoIII-like, domains remained, while the C-terminal exonuclease domain was removed from the PML alignment. Both datasets were further trimmed using trimal v1.4.rev15 with the -gappyout setting (47), resulting in a PML data set containing 545 sites and a ExoIII dataset containing 242 sites. These datasets were used to estimate maximum-likelihood phylogenies using IQTree v1.5.5 (48) using the LG + C60 + gamma model of evolution.

PML RBCC domain phylogeny

The resulting PML phylogeny (Figure 1), created using the RBCC domains of the protein, shows the expected relatedness between major groups of metazoans, with moderate support. This analysis identifies the ancestral PML originating from within Chondrichthyes (the sharks & rays). This phylogeny highlights the divergence of the PML ortholog found in *Lepisosteus oculatus,* an Actinopterygii fish identified possessing a full-length PML sequence, which, as expected, is sister to Sarcopterygii. To assess domain architectures, InterPro 88.0 (https://www.ebi.ac.uk/interpro/) was utilized. Orthologs used in the analysis include ones from the following species (corresponds to silhouettes in Figure 1): *Amblyraja radiata*, *Scyliorhinus canicular*, *Lepisosteus oculatus*, *Tachyglossus aculeatus*, *Sarcophilus harrisii*, *Mus musculus, Pteropus vampyrus*, *Capra hircus*, *Ursus maritimus*, *Homo sapiens*, *Alligator sinensis*, *Pogona vitticeps*, *Chelonia mydas* and *Coturnix japonica*.

ExoIII domain phylogeny

The phylogenetic analysis of TREX, Plex9, and PML-CDE domains (Figure 3B) shows that within metazoans, TREX and Plex proteins are distantly related and diverged before the last common ancestor of this group. Furthermore, this phylogeny shows that the ExoIII domains of PML are more closely related to Plex exonucleases. Syntenic analysis from the sequenced genomes revealed that TREX1 first appears in tetrapods due to its presence in amphibian genomes and that it likely arose from a gene duplication of TREX2. In contrast, genes encoding Plex9 sequences share synteny to the PML locus. Overall, this analysis is well supported by bootstrap values, highlighting the robust nature of the relationships uncovered, both between and within these two distinct exonuclease families.

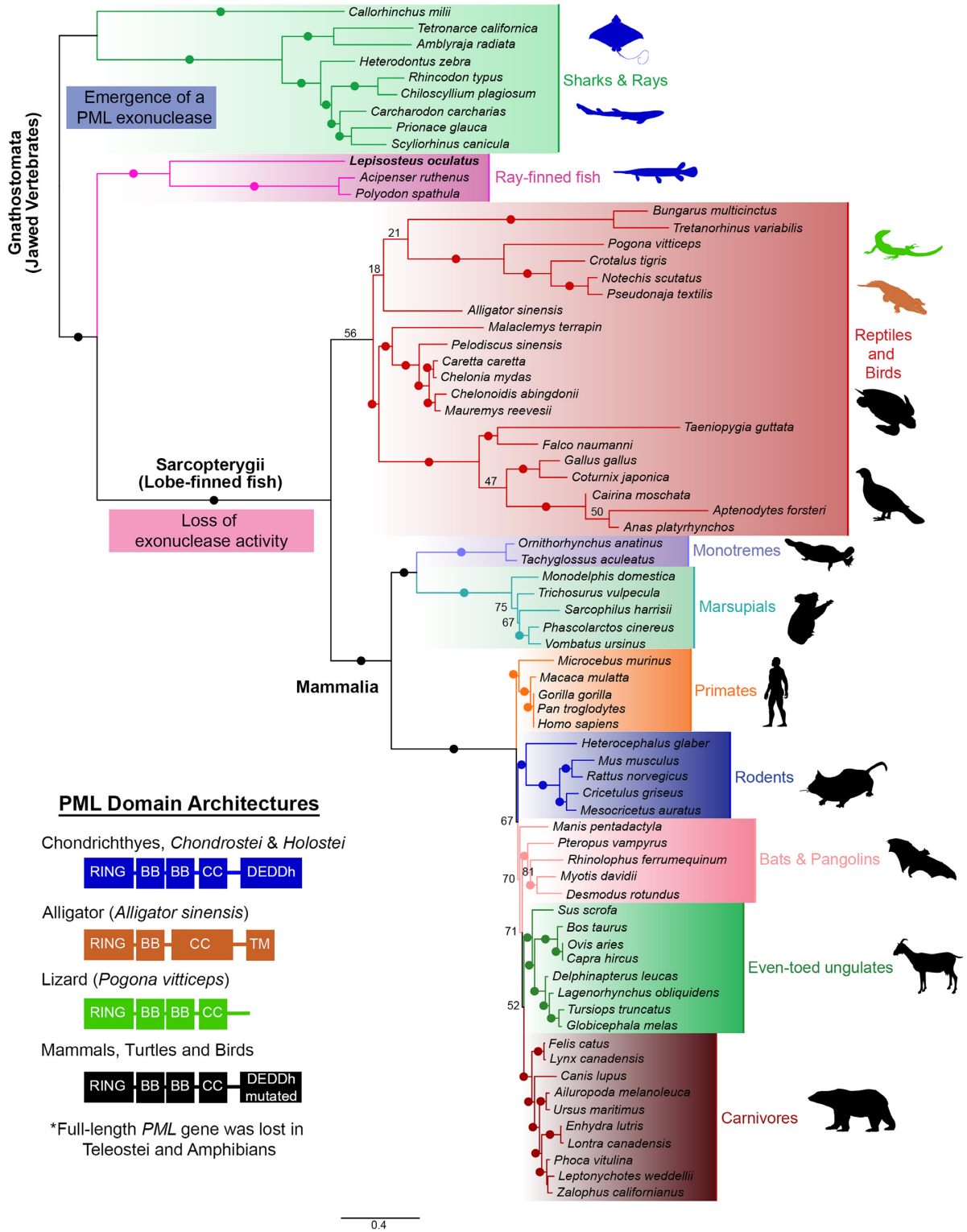

Figure 1. Domain architecture and Maximum Likelihood phylogenetic tree showing the relationships of PML homologs. 70+ protein sequences encoded in genomes from representative vertebrates were identified and utilized to construct the phylogenetic tree. PML emerged as an exonuclease in jawed vertebrates and eventually lost the catalytic residues for activity in amniote species. Teleost fishes and amphibians lack full-length PML-I. Ultrafast bootstrap values are shown as numbers if <85% and indicated with solid dots on branches if > 85%. Branch lengths indicate the expected number of amino acids substitution per site. Domains for homologs were determined using InterPro (https://www.ebi.ac.uk/interpro/) and four different domain architectures (left) were seen across analyzed sequences. Domain abbreviations refer to RBCC (Ring-B-box-Coiled Coil), BB (B-box), CC (Coiled coil), and DEDDh (3′-5′ DEDDh exonuclease). Sequences from species that were analyzed are shown in silhouettes and coloured according to the corresponding domain architecture on the left. Blue silhouettes indicate representative species with a PML ortholog encoding a predicted exonuclease domain. Silhouettes for species were obtained from PhyloPic (http://phylopic.org).

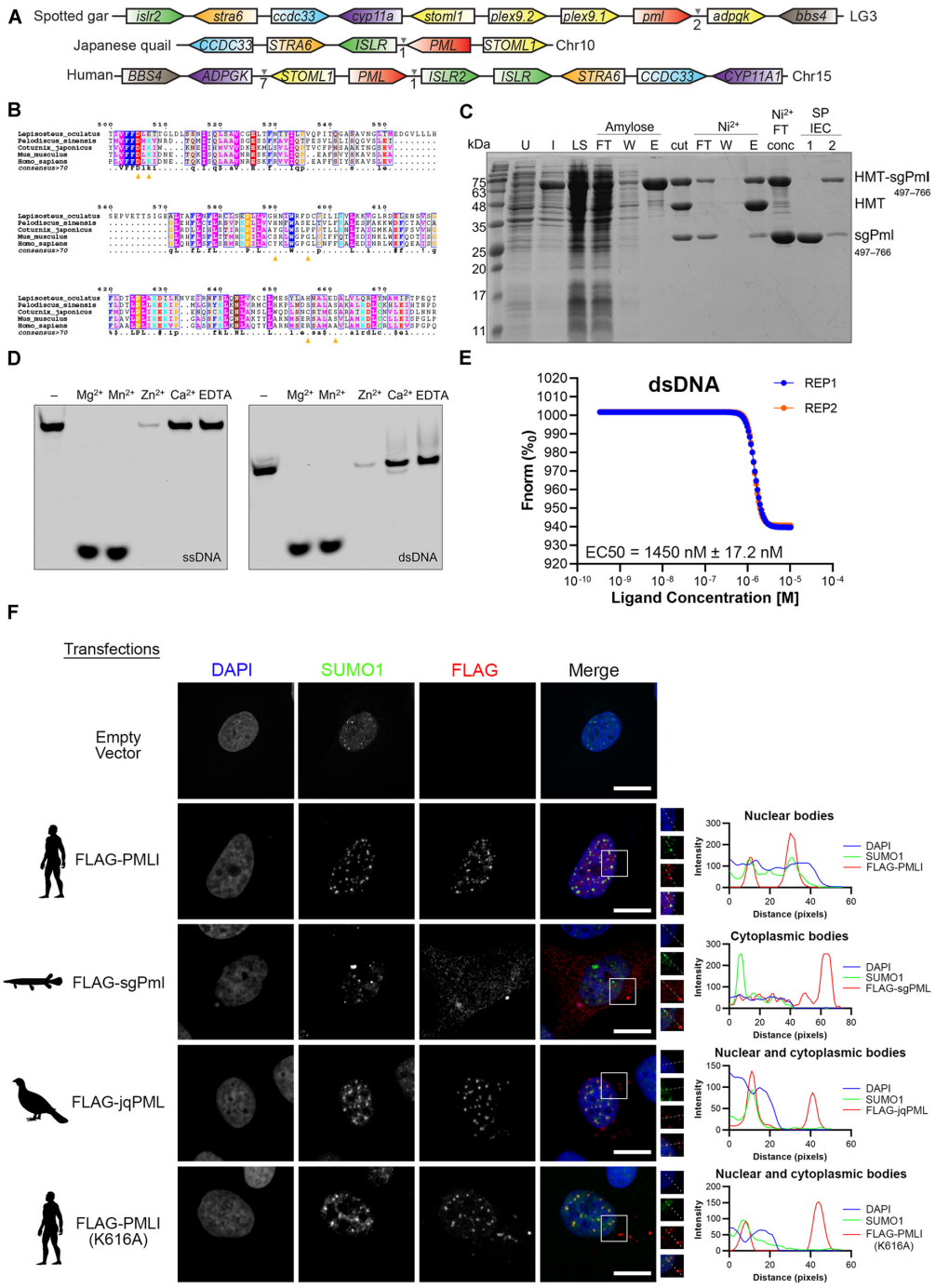

Figure 2. Spotted gar Pml is an active cytoplasmic DNA exonuclease. (**A**) Synteny of *pml* locus between the genomes of the Spotted gar, Japanese quail and Humans. Similarly coloured genes represent homologs between species. The intervening arrows with numbers indicate the number of adjacent genes between homologous genes. (**B**) MUSCLE alignment of the spotted gar PML (sgPml) C-terminal DEDDh exonuclease (CDE) domain protein with other PML orthologs. The consensus sequence is displayed below the sequences for residues with over 70% conservation across species. Yellow arrows indicate predicated catalytic residues for sgPml and the red circle indicates the conserved SUMOylation site (PML-I K616). (**C**) Purification of sgPml-CDE. sgPml-CDE was expressed as a fusion to sequences encoding for hexahistidine, maltose binding protein, and a TEV protease cleavage site (HMT). SDS-PAGE analysis is shown for samples of uninduced (U) and induced (I) cells, soluble lysate (LS), after TEV cleavage (cut), flowthrough (FT), wash (W), and elution (E) fractions of amylose and Ni^{2+} affinity chromatography, and fractions from ion exchange chromatography (IEC). (**D**) Exonuclease activity of sgPml-CDE requires Mg^{2+} or Mn^{2+}. Fluorescent oligonucleotides were incubated with sgPml-CDE in the presence of the indicated divalent cation or EDTA. '-' refers to input without the addition of cations. (**E**) sgPml binds to dsDNA with high affinity. Microscale thermophoresis was used to quantify the affinity between sgPml-CDE and dsDNA (n = 2). The different coloured lines are data shown for two replicates. (**F**) Subcellular localizations of FLAG-tagged PML-I, sgPml, quail (jqPML) and PML-I (K616A). Cells were transfected with either an empty vector or the indicated FLAG-tagged proteins. Colocalization between FLAG-tagged proteins and SUMO1 was assessed by immunostaining and shown at the right as line plot for the region of interest bound by the white box in the merged fluorescence image. Additional fields of view are shown in Supplementary Figure S16. Scale bars represent 10 μm.

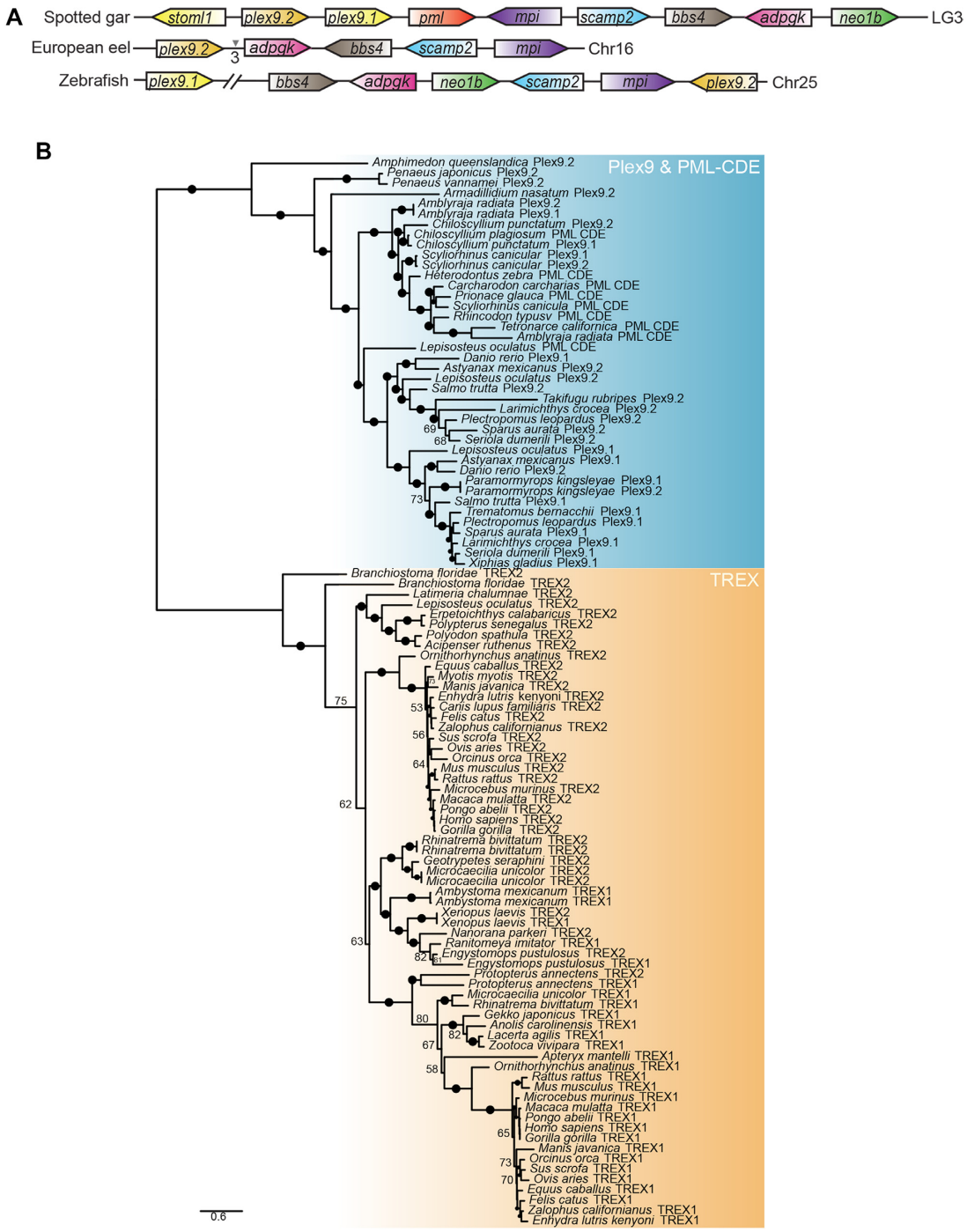

Figure 3. Origins of PML-like exon 9 (*Plex9*) and *TREX1* genes. (**A**) Teleost *plex9* genes show synteny to the spotted gar *pml* locus. Similarly coloured genes represent homologs between species. The chromosome where the *pml* locus is found in each species is indicated on the right. The intervening arrows with numbers indicate the number of intervening adjacent genes that exist between the homologous genes not found in the syntenic region. (**B**) Plex9 homologs resemble both TREX1/2 and PML exonucleases but are related to PML exonucleases. A maximum likelihood tree was constructed using Plex9, PML CDE and TREX1/2 protein sequences. While the TREX1 proteins cluster with TREX2, sequences for Plex9 cluster with PML CDE sequences. Bootstrap values are indicated beside each node. Ultrafast bootstrap values are shown as numbers if <85% and indicated with solid dots on branches if >85%. Branch lengths indicate the expected number of amino acid substitutions per site.

Plasmid construction

Synthetic genes coding for zebrafish Plex9.1 (Ile44-Leu244; Ensembl:ENSDARG00000092584; NCBI: XP_005174272.1) and spotted gar PML-CDE (Ala497–Glu766; Ensembl: ENSLOCG00000014935; NCBI: XP_015199146.1) were purchased (BioBasic Inc) and ligated into a modified pET21 expression vector that contained upstream sequences coding for a hexahistidine tag (H), maltose binding protein (M), and tobacco etch virus protease recognition sequence (T) to create pHMT-zfPlex9 and pHMT-sgPml-CDE, respectively. Site directed mutagenesis was used to create pHMT-zfPlex9.1 (D61N) using pHMT-zfPlex9.1 as a template. The fidelity of all plasmids was verified by sanger sequencing (Eurofins Genomics).

In addition, along with *zfplex9.1* and *sgpml*, another zebrafish gene coding for *plex9.2* (Ensembl ID: ENSDARG00000093773; NCBI: XP_002666829.1) was utilized in the study. In addition, axolotl TREX1/2/3 sequences (AMEXTC_0340000192613, AMEXTC_0340000056847, AMEXTC_0340000061536) from the axolotl transcriptome Am_3.4 available on Axolotl-omics (https://www.axolotl-omics.org). These coding gene sequences were all cloned into CMV-based FLAG-tagging and Clover-tagging vectors (CMV-FLAG-J1 and CMV-Clover-J1; derived from commercially available vectors pEGFP-C1 and pEGFP-N1; Clontech) for ectopic expression in human cells.

RNA used for cloning the axolotl *TREX*, spotted gar *pml*, and zebrafish *plex9.1* and *plex9.2* sequences was obtained from animal tissues (described below under animal ethics). Two gar individuals were raised as described in Darnet et al. (2019) to a total length of ∼30 cm, euthanized with a lethal dose of MS-222 (Sigma); then tissues (muscle, gills, intestine, liver, brain, spleen, heart, gas bladder) were dissected, and stored in RNAlater (Ambion) (49). A caudal fin clip was stored in 95% ethanol. From the available tissue, expression of *sgpml* was highest in the spleen and thus this was utilized for downstream cloning. RNA for downstream cloning was derived from axolotl limbs that were dissected and stored in Trizol. In the case of zebrafish, the RNA is derived from 24 hpf zebrafish embryos. Quail RNA for quail PML cDNA was isolated from a pellet of QM5 cells (quail muscle five cells). The RNeasy (Qiagen) kit was used for all extractions, according to the manufacturer's instructions. The cDNAs were prepared using the SuperScript™ IV One-Step RT-PCR System (Invitrogen) and sequenced prior to subcloning into indicated vectors.

For *Polyodon spathula* (NCBI: XP_041075354) and *Terrapene carolina triunguis* (Ensembl ID: ENSTMTG00000003177, NCBI: XP_024070906.1) *PML* coding gene sequence DNA was synthesized through gBlocks™ Gene Fragments in two parts (Integrated DNA Technologies). Overlap extension PCR was utilized to generate the full sequence, which was then cloned into the CMV driven FLAG-tag vectors used for the other coding sequences. Human PML-I K616A was generated using the Q5® Site-Directed Mutagenesis Kit (E0554S) and cloned with the CMV driven FLAG-tag vector.

Protein expression and purification of zfPlex9.1 and sgPml-CDE

BL21(DE3) *E. coli* (New England BioLabs Inc.) were transformed with pHMT-zfPlex9.1, pHMT-zfPlex9.1 (D61N) or pHMT-sgPml-CDE. Transformed cells were grown in LB media to an OD_{600} of ∼0.6, after which protein expression was induced with 0.5 mM IPTG. After overnight expression at 20°C (16°C for pHMT-zfPlex9.1 (D16N)), cells were resuspended in lysis buffer (HMT-ZfPLEX9.1: 20 mM Tris pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol (BME); HMT-ZfPLEX9.1 (D61N): 20 mM CHES pH 9, 100 mM NaCl, 5 mM BME, 5 mM $MgCl_2$; HMT-sgPml-CDE: 20 mM Tris pH 8, 500 mM NaCl, 5 mM BME), lysed by sonication, and clarified by centrifugation (25 000 × g for 20 min at 4°C). The supernatant was loaded onto an amylose chromatography column (New England BioLabs Inc.), washed with lysis buffer, and eluted with lysis buffer containing 10 mM maltose. The proteins were then incubated with TEV protease overnight in elution buffer (HMT-ZfPLEX9.1) or with concurrent overnight dialysis at 4°C against dialysis buffer (HMT-ZfPLEX9.1 (D61N): 20 mM CHES pH 9, 100 mM NaCl, 5 mM BME, 5 mM $MgCl_2$; HMT-sgPml-CDE: 20 mM Tris pH 8, 250 mM NaCl, 5 mM BME). After cleavage, exonuclease domains were isolated using Ni^{2+} affinity chromatography and then purified by size exclusion chromatography (HiLoadTM 16/600 SuperdexTM 75 pg) using 20 mM Tris pH 7.5, 100 mM NaCl, 5 mM BME (HMT-ZfPLEX9.1 and HMT-ZfPLEX9.1 (D61N)) or diluted 5-fold with 20 mM MES (2-(*N*-morpholino) ethanesulfonic acid) pH 6, 5 mM BME and purified by cation exchange chromatography (HiTrap™ SP HP, cytiva). Purifications were monitored by SDS-PAGE and UV/Vis absorbance (using predicted extinction coefficients at 280 nm of 20 970, 20 970 and 18 450 $M^{-1}cm^{-1}$ for ZfPLEX9.1, ZfPLEX9.1 (D61N), and sgPml-CDE, respectively).

In vitro exonuclease assays

Unless otherwise noted, exonuclease reactions (40 μl) contained 500 nM 5′-6-FAM labeled oligonucleotide (ssDNA: 5′-ATACGACGGTGACAGTGTTGTCAGACAG GT-3′or dsDNA pseudo-palindrome: 5′-TCACGTGCTG AC/GTCAGCACGACG-3′), 20 mM Tris pH 7.5, 2 mM dithiothreitol, and 100 μg/ml BSA. For metal specificity assays the reactions contained 625 nM of zfPlex9.1 or sgPml-CDE and 2 mM metal ($MgCl_2$, $MnCl_2$, $ZnSO_4$, or $CaCl_2$) or 10 mM EDTA. Exonuclease titration assays contained 4.9–625 nM of zfPlex9.1 or zfPlex9.1 (D61N) and 5 mM $MgCl_2$. Reactions were allowed to proceed for 20 min at room temperature and then quenched with 3 volumes of 100% ethanol, dried via speed vac, resuspended in 10 μl formamide, resolved by urea PAGE (19% 29:1 acrylamide and 7 M urea in TBE), and visualized using a fluorescence imager (VersaDoc).

Microscale thermophoresis (MST) binding assays

Binding assays were carried out in 20 mM Tris pH 7.5, 100 mM NaCl, 5 mM $CaCl_2$ and 5 mM BME. Solutions contained 50 nM ssDNA or dsDNA and zfPlex9.1, zfPlex9.1

(D61N), or sgPml-CDE to maximum concentrations of 11.7, 15.2 and 11.1 μM, respectively. Thermophoresis measurements were collected at 25°C using a Monolith NT.115 (NanoTemper Technologies) set to medium LED power and 20% excitation power. Raw data were processed using NT Analysis software (NanoTemper Technologies) and fit to a Hill binding model to determine the EC_{50} and Hill coefficient.

Cell culture

U2OS osteosarcoma cell lines (parental U2OS, $U2OS^{Clover–PML}$ (50), $U2OS^{GFP–PML-I}$, U2OS TREX1 KO cells, U2OS PML KO cells (13)) were cultured in Dulbecco's modified Eagle's medium (Life Technologies) supplemented with 10% fetal calf serum, at 37°C with 5% CO_2. QM5 (Quail Muscle clone 5) cells were a gift from the Roy Duncan lab (Dalhousie University) and grown like U2OS. NHDF (WT and PML KO cells) were cultured in alpha-modified Minimum Essential Medium (Life Technologies) supplemented with 15% fetal calf serum and GlutaMAX (Thermo Fisher) at 37°C with 5% CO_2.

Generation of a CRISPR/cas9 KO lines

PML CRISPR/Cas9 KO lines, both immortalized NHDF and U2OS, were previously generated and characterized (13). We generated U2OS cells lacking *TREX1* expression, we designed a strategy to knock-in a puromycin resistance cassette into exon 2 of *TREX1*, thereby disrupting *TREX1* gene expression but not overlapping with the *ATRIP* protein encoding region. U2OS was used a cell line since they have impaired cGAS-STING signalling and allow for us to focus on L1-retrotransposition in an interferon-independent manner (51).

Two guide RNAs targeting the Cas9D10A nuclease to the *TREX1* locus were designed utilizing the CHOPCHOP v3 webtool for designing guide RNAs (52). We selected the top two guide RNAs, g1 (5′-CCCAACCATGGGCTCGCAGGCCCTGCCCCCGG GGCCCATGCAGACCCTCATCTTTTTCGACATGG-3′) and g2 (5′- CCATGTATGGGGTCACAGCCTCTG CTAGGACCAAGCCAAGACCATCTGCTGTCACA ACCACTGCACACCTGG-3′) and cloned these guides into the All-in-one (AIO)-PURO vector encoding the Cas9D10A nickase (Addgene #74630). U2OS cells were treated with 2 μg/ml of puromycin (Invitrogen) and resultant clones were isolated. Resultant clones were screened by both western blotting and immunofluorescence to confirm that TREX1 expression was absent (Supplementary Figure S12).

We also examined *ATRIP* expression in the *TREX1* KO lines, which was previously ignored in other reports of a human *TREX1* KO line, despite *TREX1* being nested in the locus of *ATRIP* (53,54). We were unable to get a positive clone with guide 2 (g2) and this appears to be a result of ATRIP levels being altered, where two prominent bands are present (data not shown). We therefore utilized the clone generated from guide 1 (which still show a faint secondary ATRIP protein band that is not present in WT cells) (Supplementary Figure S12). We utilized this clone as TREX KO#1 for experiments involving L1 suppression, a previously established function for TREX1 (both via KO models in mice and overexpression studies) (55,56). Our addback of TREX1 restored L1 suppression, supporting the notion that elevated L1 retrotransposition was a result of TREX1 loss and not ATRIP being altered (Figure 5).

LINE retrotransposition assays

To assess the effect of proteins on retrotransposition, we utilized methods for previously characterized plasmids encoding a LINE retroelement (human L1, zfL2-1, zfL2-2, UnaL1 or human L1 with T7-tagged ORF1p) that encode a neomycin cassette that when transfected into cells, is only successfully re-integrated upon retrotransposition (32,35,57–61). Upon G418 (Thermo Fisher) selection, it is possible to quantify the number of positive clones that survived which will be proportional to the LINE activity occurring. Co-expression of the proteins of interest with the LINE encoding plasmid, will determine which proteins enhance, suppress, or have no impact on cell LINE retrotransposition. HeLa and U2OS cells were utilized for LINE retrotransposition since they are susceptible to high levels of LINE retrotransposition in culture (32,57).

Cells (HeLa or U2OS) were seeded in a 6-well at densities of 2×10^5 per well, left to attach for 24 h and co-transfected using Lipofectamine 3000 (Roche) with (1) 1 ug of a retrotransposon encoding plasmid (2.5 ug for the zfL2-1 retrotransposon) and (2) with 500 ng of a FLAG-tagged protein of interest. The only exception was the GFP-TREX1(D18N) plasmid which was obtained from Addgene (27220). No difference was observed between cell toxicity or transfection efficiencies between GFP-TREX1(D18N) and the other plasmids (Supplementary Figures S1 and S2). After 2 days post-transfection, the cells were reseeded on a 10 cm plate, allowed to adhere for 24 h, and then incubated with G418 (500 ug/ml for HeLa; 400 ug/ml for U2OS) for a total of 10 days for HeLa and 14 days for U2OS. Then post-selection, the resistant colonies were fixed with 1 ml of methanol for 20 min. The colonies were then washed with PBS and stained with 0.5% crystal violet (stained for 30 min in 5% acetic acid and 2.5% isopropanol) for 30 min. The entire fixation and staining procedure were completed at room temperature. Colonies were then counted and retrotransposition activity was determined.

To determine protein levels of T7-tagged ORF1p after ubiquitin-proteasome inhibition. Cells on a 10 cm^2 plate were transfected with 5 μg of the L1 plasmid encoding T7-ORF1p or an empty vector using Lipofectamine 2000. For addback experiments, cells were transfected with an additional 5 μg of FLAG-PML-I, FLAG-PML-IV, or FLAG empty vectors ('-' control). Then, after 24 h, we treated U2OS cells with MG132 (Sigma; M7449) at varying concentrations (1, 2.5, 5 μM) for 8 h as described previously (62). Cells were then washed with PBS twice and samples were handled for western blotting (described below).

2′,3′-cGAMP quantification

4×10^6 or 6×10^6 U2OS cells (for TREX1 KO and PML KO experiments, respectively) were seeded into 15-cm

dishes, and 24 h later cells were transfected with the Bluescript vector (empty vector control) and Human L1 plasmid (used in L1 retrotransposition assay) using Lipofectamine 2000 reagent (Invitrogen). Cells were harvested 36 h after transfection, washed with 2× with PBS and pelleted before lysis. Samples were resuspended in 500 μl M-PER (mammalian protein extraction reagent) lysis buffer (Thermo Scientific). Lysates were incubated on ice for 30 min with gentle agitation every 10 min, before being spun down at 16 000 × g, 4°C for 10 min. Samples were quantified using the 2′3′cGAMP ELISA Kit (Cayman Chemical) according to the manufacturer's instructions.

Immunofluorescence microscopy

For transfections, cells were seeded into wells containing coverslips in 6-well dishes, then transfected the next day with expression vectors. One day after transfection, coverslips were washed briefly with PBS and the cells were fixed in 2% PFA, permeabilized with 0.5% Triton X-100 in PBS and blocked with 4% BSA in PBS. Cells were then immunolabeled with primary antibodies specific for FLAG (mouse anti-FLAG M2, Sigma, F3165, 1:200), T7 (rabbit anti-T7, Millipore/Sigma, AB3790, 1:200), PML (two antibodies used; sheep anti-PML, Diagnostics Scotland, PML2A, 1:500; rabbit anti-PML, Bethyl Laboratories, A301-167A, 1:1000), mouse anti-T7 (Millipore, AB3790, 1:200), mouse anti-FLAG (Millipore, F1804, 1:500), SUMO (two antibodies used; mouse anti-SUMO1, clone 21C7, Zymed, #33-2400, 1:200; rabbit anti-SUMO1, Abcam, ab32058, 1:200), rabbit anti-DAXX (polyclonal D7810; Sigma-Aldrich, 1:500), and TREX1 (rabbit anti-TREX1, Abcam, ab185228, 1:400). Then for secondary staining, coverslips were washed with PBS and incubated with Alexa-Fluor 647 donkey anti-rabbit, Alexa-Fluor 488 donkey anti-sheep, Alexa-Fluor 488 donkey anti-rabbit, and Alexa-Fluor 555 donkey anti-mouse (Thermo Fisher Scientific) secondary antibodies. Finally, the cells were washed several times in PBS and incubated with 1 μg/ml of 4',6-diamidino-2-phenylindole (DAPI) (Sigma) to visualize the nuclei.

For transfections, we marked ER and L1 RNP structures using plasmids obtained from Addgene and individuals. The ER was marked using a Cytochrome p450 (CytERM) tagged to mScarlet (Addgene #85066), which was utilized as an ER marker. T7-tagged ORF1p was previously used to mark L1 RNPs and encoded in the pES2TE1 vector. Plasmids for flag-tagging of proteins of interest are described above under plasmid construction, aside from FLAG-TREX1 (Addgene #27218), GFP-TREX1 (Addgene #27219) and GFP-TREX1 (D18N) (Addgene #27220) which were obtained from Addgene. U2OS cells were transfected using Lipofectamine 2000 (Invitrogen) and HeLa cells using Lipofectamine 3000 (Invitrogen), according to the manufacturer instructions.

Fluorescent micrographs were captured with a Prime95 scientific complementary metal-oxide-semiconductor (sCMOS) camera (Photometrics) using a custom-built spinning-disk confocal laser Zeiss Cell Observer Microscope (Intelligent Imaging Innovations, 3i) equipped with a 1.4 NA 63× immersion oil objective lens and both laser (3i) and LED illumination via a Spectra light engine (Lumencor). Images were captured and processed using Slidebook 6.1 (3i) and assembled into figures using Adobe Photoshop CS5.

Western blotting

For western blot analysis of SUMOylation status, cells were recovered from confluent 10 cm culture dishes and washed with PBS. U2OS cells were treated for 24 h with 1 μM ML-792 (Selleckchem), a SUMO E1 inhibitor (targets SAE1) to inhibit total SUMOylation within cells (63). After treatment, cells were then washed twice with PBS and either harvested for western blotting or fixed for immunofluorescence staining. Cells were lysed on ice for 20 min in 6 mM urea lysis buffer with protease inhibitors (P8340, Sigma) and sonicated, before further processing. For all other samples, the cells were lysed on ice for 20 min in RIPA buffer (Sigma) with protease inhibitors (P8340, Sigma). Lysates were cleared (10 min, 15 000 × g, 4°C) and protein extracts were analyzed by SDS-PAGE and western blotting using 5% milk powder with 0.1% Tween 20 in PBS as a blocking solution. Protein was transferred to nitrocellulose membrane (BioRad). Total protein was determined directly on the membrane using the 4–15% Mini-PROTEAN TGX-Stain Free Protein Gel system (BioRad). For homemade gels, we used actin as our loading control. Antibodies used for Western blotting analysis were: rabbit anti-SUMO1 (Abcam, ab32058, 1:2000), rabbit anti-ubiquitin (linkage K48) (Abcam, ab140601, 1:4000), mouse anti-T7 (Millipore, AB3790, 1:2000), mouse anti-FLAG (Millipore, F1804, 1:2000), mouse anti-actin (Sigma, A2228, 1:5000), rabbit anti-TREX1 (Abcam, ab185228, 1:1000) and anti-ATRIP (Sigma, PLA1030, 1:1000). We have provided uncropped, original image files for blots used in the figures within the supplemental material (Supplementary Figures S17–S19).

Statistical analysis

All statistical analyses were performed using GraphPad Prism 9.0.1. The sample size and error bars for each experiment are defined in the figure legends. Comparisons between groups for LINE1 retrotransposition and the 2′3′-cGAMP assay were analysed by a repeated measures one-way ANOVA, with the Geisser-Greenhouse correction and then a Tukey's multiple comparisons test between groups.

Animal ethics

Spotted gar work was approved by the Institutional Animal Care and Use Committee at Michigan State University (protocol no. AUF 10/16-179-00). For axolotl work, Axolotl Université de Montreal animal ethics committee authorization was: 20-103. The Université de Montral animal ethics committee is recognized by the CCAC. The use of zebrafish embryos for generation of RNA was approved by Dalhousie University Committee on Laboratory Animals (Protocol Number 20-130)

RESULTS

PML first emerged in jawed vertebrates

To elucidate the evolutionary history of *PML*, we identified putative orthologs in publicly available databases of eukaryote genomes and transcriptomes (Figure 1). PML-I orthologs were identified in jawed vertebrates (Gnathostomata) including extant cartilaginous fishes (Chondrichthyes) and bony vertebrates (Euteleostomi) (64). We could not identify orthologs of PML in agnathan cyclostome (lamprey and hagfish) genomes (Figure 1). In addition, we documented the divergence of the *PML* gene in amniotes (Figure 1).

Among Chondrichthyes, PML ortholog architecture can be stratified into three groups, those with genomes encoding full length PML-I, those encoding putative orthologs with only a conserved RBCC domain, and those with a C-terminal transmembrane domain (Figure 1). While the RBCC domain is common among TRIM family proteins (65), the C-terminal domain of PML-I encoded by exon 9 uniquely shares homology with DEDDh exonucleases, which was originally annotated as a putative exonuclease III domain (42). This C-terminal DEDDh exonuclease (CDE) domain is conserved in bird, reptile, and mammalian PML orthologs as a predicted enzyme fold; however, the predicted catalytic residues are mutated for these orthologs of PML (Figure 1). Uniquely in ray-finned fish (such as paddlefish, spotted gar and sturgeon) and in cartilaginous fishes such as sharks and rays, PML orthologs have a CDE with intact catalytic residues, suggesting that they could be functional exonucleases (Figure 1).

During vertebrate evolution, the PML gene appears to have been lost several times in major extant euteleostome lineages, including amphibians and teleost fishes (Figure 1). We also failed to identify a PML ortholog in the coelacanth genome, the most basally diverging lineage of extant Sarcopterygii (lobe-finned fishes). Similarly, we could not identify an ortholog of PML in the lungfish genome. Within reptiles, turtles show conservation of the CDE, whereas lizard genomes encode a truncated PML ortholog lacking the CDE. Birds similarly have PML homologs that resemble full-length PML. Since turtles are thought to be more closely related to archosaurs than lepidosaurs, our analyses suggest that bird and turtle ancestral lineages retained an ortholog resembling a full-length PML-I isoform (66).

PML was once a cytoplasmic exonuclease

The discovery of full length PML-I orthologs in ray-finned fish, sharks and rays encoding putatively active CDE domains was intriguing. We utilized the spotted gar genome to further explore PML-I from ray-finned fish (67). The spotted gar lineage diverged ∼350 million years ago from the teleost lineage before the teleost genome duplication event and has a slowly evolving genome (67). These features make the spotted gar an excellent representative species for studying the *PML* gene in fish. At the locus encoding the spotted gar *pml* ortholog (referred to as *sgpml*), there is strong synteny to the human *PML* locus on human chromosome 15, with flanking genes *STOML1, ISLR, CCDC33* being conserved within the syntenic region (Figure 2A). We also utilized a bridging amniote species, *Coturnix japonica* (Japanese quail) that encodes a *PML* locus (referred to as *jqPML*) with synteny to both the spotted gar and human loci (Figure 2A).

Given that the sgPml protein encodes a potentially active CDE (Figure 2B), we cloned the full length *sgPml* gene from gar splenic tissue for further characterization. We also expressed the codon optimized putative DEDDh exonuclease of sgPml (residues 497–766) and purified it to homogeneity (Figure 2C). We found that the sgPml-CDE encodes an active exonuclease capable of degrading ssDNA and dsDNA when incubated with Mg^{2+} and Mn^{2+} (Figure 2D). Thus, akin to other DEDDh family exonucleases (68), it appears that sgPml can degrade dsDNA in the presence of divalent cations. In addition, using microscale thermophoresis binding assays in the presence of Ca^{2+}, we found that sgPml-CDE can bind to dsDNA with an EC50 of 1450 nM ± 17.2 nM (Figure 2E) but bound ssDNA at low affinity. These data indicate that sgPml has the capacity to bind and degrade both ssDNA and dsDNA but with a higher affinity for dsDNA through its CDE.

Nuclear PML bodies emerged in amniotes

We next characterized the cellular localization of sgPml by expressing it in human wild type (WT) and PML KO U2OS osteosarcoma cells. Strikingly, in contrast to the mammalian PML protein, sgPml did not form nuclear bodies and localized diffusely in the cytoplasm, forming ∼1–3 cytoplasmic body-like puncta (Figure 2F). We also expressed the American Paddlefish PML ortholog, another predicted PML exonuclease and observed cytoplasmic localization like sgPml (Supplementary Figure S3A). Since mammalian PML NBs are associated with post-translation modifications of nuclear body proteins by the small ubiquitin like modifier (SUMO) (66), we also examined SUMO1 localization relative to these cytoplasmic sgPml puncta. These sgPml puncta were negative for SUMO1 immuno-staining, unlike mammalian PML nuclear bodies that accumulate SUMO1. To assess whether sgPml formed nuclear bodies in the absence of SUMOylation, we treated cells with ML-792, a SUMO E1 inhibitor that blocks total SUMOylation within the cell (Supplementary Figure S4A). We found that sgPml was still capable of forming cytoplasmic puncta even after SAE1 inhibitor treatment, consistent with these cytoplasmic sgPml bodies forming in a SUMO-independent manner (Supplementary Figure S4A).

Sequence comparisons revealed that the target lysines involved in human PML SUMOylation, which first appear in amniote species according to our evolutionary reconstruction, align partly to the CDE and its catalytic residues (Figure 2B). This could suggest that a shift in PML cellular function and localization may have occurred concurrently with the acquisition of SUMOylation. To assess the possibility that SUMO site evolution drove the nuclear localization and function of PML, we cloned and expressed the Japanese quail PML ortholog (jqPML) (Figure 2B). The expression of jqPML and human PML-I in quail cells all led to the formation of SUMO1 colocalizing nuclear bodies (Supplementary Figure S3B). Similarly, expression of jqPML in human cells also led to SUMO1 accumulating at jqPML

nuclear bodies (Figure 2F). However, like sgPml, jqPML also formed cytoplasmic bodies lacking SUMO1 (Figure 2F, Supplemental Figure S16). The expression of a PML ortholog from a turtle species mirrored the localization of jqPML (Supplementary Figure S3A).

Human PML K616 is one of the lysine residues with the potential to be poly-SUMOylated (69). Intriguingly, the residue also aligns to a catalytic residue of sgPml. The conserved lysine also appears to have been one of the first potential SUMOylation sites appearing in amniote orthologs of PML, suggesting that it may have contributed to the shift in PML localization from cytoplasm to nucleus (Figure 2B). We generated a PML-I K616A mutant and found that its localization shifted to the cytoplasm, where it formed numerous cytoplasmic puncta that were negative for SUMO1 (Figure 2F). Human PML-I K616A is likely able to still maintain some protein at nuclear bodies owing to its other SUMOylation sites. The localization resembled that of jqPML, which both formed nuclear bodies and cytoplasmic puncta (Figure 2F). Thus, the PML-I K616 residue appears to play an important and conserved role across species in dictating the nuclear localization of PML.

When the ectopic expression of orthologs was examined by western blotting, FLAG-jqPML appeared as multiple bands like human PML-I (Supplementary Figure S4B). However, FLAG-sgPml appeared as a single band indicating that it is not likely modified by SUMO as suggested by the immuno-staining (Supplementary Figure S4B). The jqPML higher molecular weight band migrated approximately ∼12 kDa higher, which would be the predicted molecular weight of jqPML conjugated to SUMO1 (a single moiety). Treatment of cells with the SAE1 inhibitor ML-792, thus blocking global SUMOylation, resulted in a reduction in the higher molecular weight jqPML and human PML bands (Supplementary Figure S4B). These results support the importance of PML SUMOylation for nuclear body formation, and indicates that avian and turtle PML orthologs, with the acquisition of SUMO-target lysine residues, can localize to both the cytoplasm and the nucleus to form PML NBs.

Plex9 proteins are novel DEDDh exonucleases

Most ray-finned fish species (99.8%) are teleost fish that underwent a genome duplication event accompanied by subsequent gene order rearrangements (70). During this event, it appears that the full length *PML* gene was lost in teleost fishes. However, we identified a pair of genes sharing a high degree of homology to the CDE of sgPml in zebrafish. We describe these novel genes as Pml-like exon 9 genes (*Plex9.1* and *Plex9.2*), which like the PML-CDE, share sequence similarity with TREX1 and TREX2, the two major DEDDh exonucleases in mammals (Supplementary Figure S5A) (68,71). A closer examination of the spotted gar *pml* locus also revealed the presence of two adjacent upstream genes that appear to have duplicated from the PML-CDE (Figure 3A). Phylogenetic analyses also indicated that Teleostei Plex9 shares the highest similarity to the PML-CDE, and across genera, TREX1 orthologs consistently cluster independently from Plex9 orthologs (Figure 3B). Thus, *plex9* genes in Telostei are DEDDh exonucleases closely related to the CDE of PML-I and are not evolutionarily related to *TREX1/2*.

We also identified Plex9 orthologs in other ray-finned fishes and in the elasmobranch sub-class of cartilaginous fish, whereas the TREX1 gene first appears in the tetrapod ancestor since it is found in amphibian species and amniotes (Figure 3B). Neither *plex9.1* or *plex9.2* orthologs can be identified in lobe-finned fish (Figure 3B). TREX2 predates TREX1 and can be found conserved in cephalochordates such as *Branchiostoma floridae*. Syntenic analysis of the *TREX1* loci reveals that *TREX1* was likely derived from a duplication of the *TREX2* gene (Supplementary Figure S5B). Thus, while tetrapods lost the *Plex9* genes, they retained the newly emerged *TREX1* gene and the *PML* gene encoding a CDE domain. In amphibians, we identified a putative ortholog of *TREX1* but could not identify genes resembling *PML* or the *plex9* orthologs (Supplementary Figure S5B). The axolotl genome encodes a TREX1 ortholog and there is expression of three transcripts for TREX1-like proteins (*axTREX1*, *axTREX2*, *axTREX3*). Previously, it was thought that *TREX1* was exclusive to mammals (72), however it appears that the progenitor *TREX1* gene encoding a DEDDh exonuclease is more generally present in tetrapod genomes.

To better understand the function of the Plex9 and non-mammalian TREX1 proteins, we cloned the zebrafish *plex9* genes (*zfplex9.1* and *zfplex9.2*), as well as the TREX1 ortholog isoforms from the axolotl genome (*axTREX1*, *axTREX2*, *axTREX3*) to assess their cellular roles and localization in mammalian cell assays (Figure 4A, B and Supplementary Figure S6). Expression of these orthologs in U2OS cells revealed that zfPlex9.1 as well as axTREX1 and axTREX2 localized to the ER and cytoplasm in a similar manner as TREX1 (Figure 4A), whereas zfPlex9.2 localized primarily to the cytoplasm (Figure 4A). Like TREX1 which has been linked to nucleotide excision repair and localizes to the nucleus after DNA damage (73–76), zfPlex9.1 translocated to the nucleus in response to etoposide-induced DNA damage (Supplementary Figure S7). Thus, the cellular localization of these different DEDDh exonucleases resembles that of human TREX1 (Figure 4A).

An important role for DNA exonuclease function for TREX1 is the clearance of micronuclei (77). To determine whether zfPlex9.1 and axTREX1 proteins play a role in micronuclei clearance akin to TREX1, we treated U2OS cells with reversine, which induces micronuclei by causing the mis-segregation of chromosomes (56). We found that like TREX1, zfPlex9.1 and axTREX1 also accumulated at micronuclei (Supplementary Figure S8A). In addition, we found that cells overexpressing TREX1, zfPlex9.1 and axTREX1, had less micronuclei after 48 h of 5 μM reversine treatment (Supplementary Figure S8B). Thus, these data indicate that the zfPlex9.1 and axTREX1 exonucleases have the capacity to contribute to micronuclei clearance.

We next assessed the exonuclease activity of recombinant zfPlex9.1 (aa 44–244) and found it could degrade ssDNA and dsDNA (Figure 4C, D) in the presence of Mg^{2+} and Mn^{2+}. It also bound to ssDNA (EC_{50} = 185 ± 21.8 nM) and dsDNA (EC_{50} = 180 ± 4.66 nM) with higher affinity than sgPml (Figure 4E, F). An alignment of zfPlex9.1, zfPlex9.2 and TREX1 revealed a conservation of the DEDDh

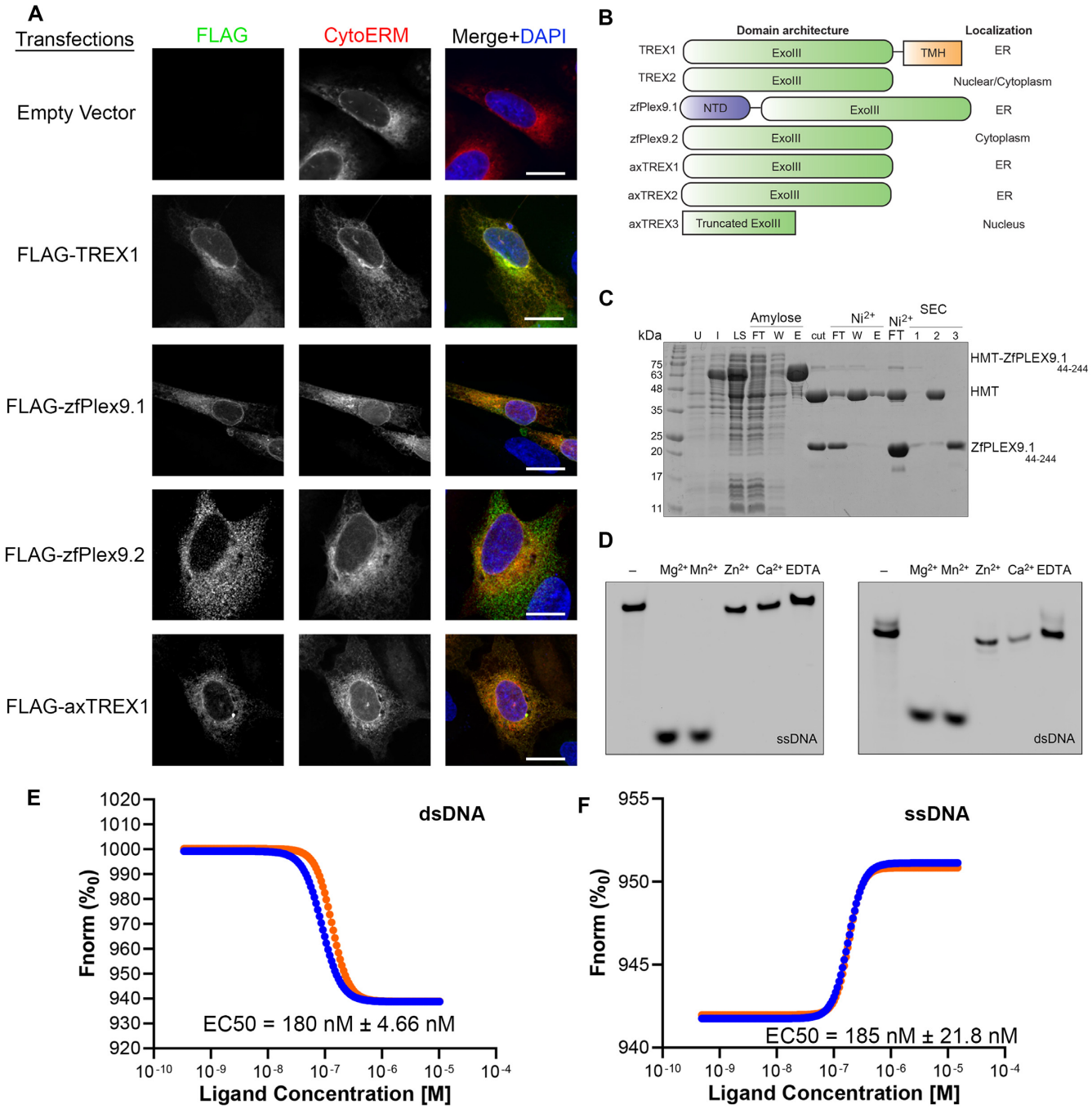

Figure 4. Plex9 proteins represent a novel class of DEDDh DNA exonucleases. (**A**) Ectopic expression of zebrafish Plex9.1/.2 (zfPlex9.1/.2) and axTREX1/2 reveals that these proteins localize to the ER and cytoplasm. Proteins were FLAG-tagged and expressed in U2OS cells to assess their localization. A domain of Cytochrome p450 (CytoERM) tagged to mScarlet (addgene #85066) was utilized as an ER marker. An empty vector (BlueScript) was used as a control. Scale bars represent 10 μm. (**B**) The different domain architectures of TREX1/2 and Plex9 proteins and their known/determined localizations are shown on the right. TREX2 localization was assessed previously as both nuclear and cytoplasmic. (**C**) Purification of zfPlex9.1. zfPlex9.1 was expressed as a fusion to sequences encoding for hexahistidine, maltose binding protein, and a TEV protease cleavage site (HMT). SDS-PAGE analysis is shown for samples of uninduced (U) and induced (I) cells, soluble lysate (LS), after TEV cleavage (cut), flowthrough (FT), wash (W), and elution (E) fractions of amylose and Ni^{2+} affinity chromatography, and fractions from size exclusion chromatography (SEC). (**D**) zfPlex9.1 requires Mg^{2+} or Mn^{2+} for activity. Fluorescent oligonucleotides were incubated with zfPlex9 and the indicated metal. Samples were resolved by UREA-PAGE and visualized via fluorescence imaging. '-' refers to input without the addition of cations. (**E**, **F**) DNA binding of zfPlex9.1. Microscale thermophoresis was used to quantify the affinity between zfPlex9.1 with (E) ssDNA and (F) dsDNA ($n = 2$). The different coloured lines are data shown for two replicates.

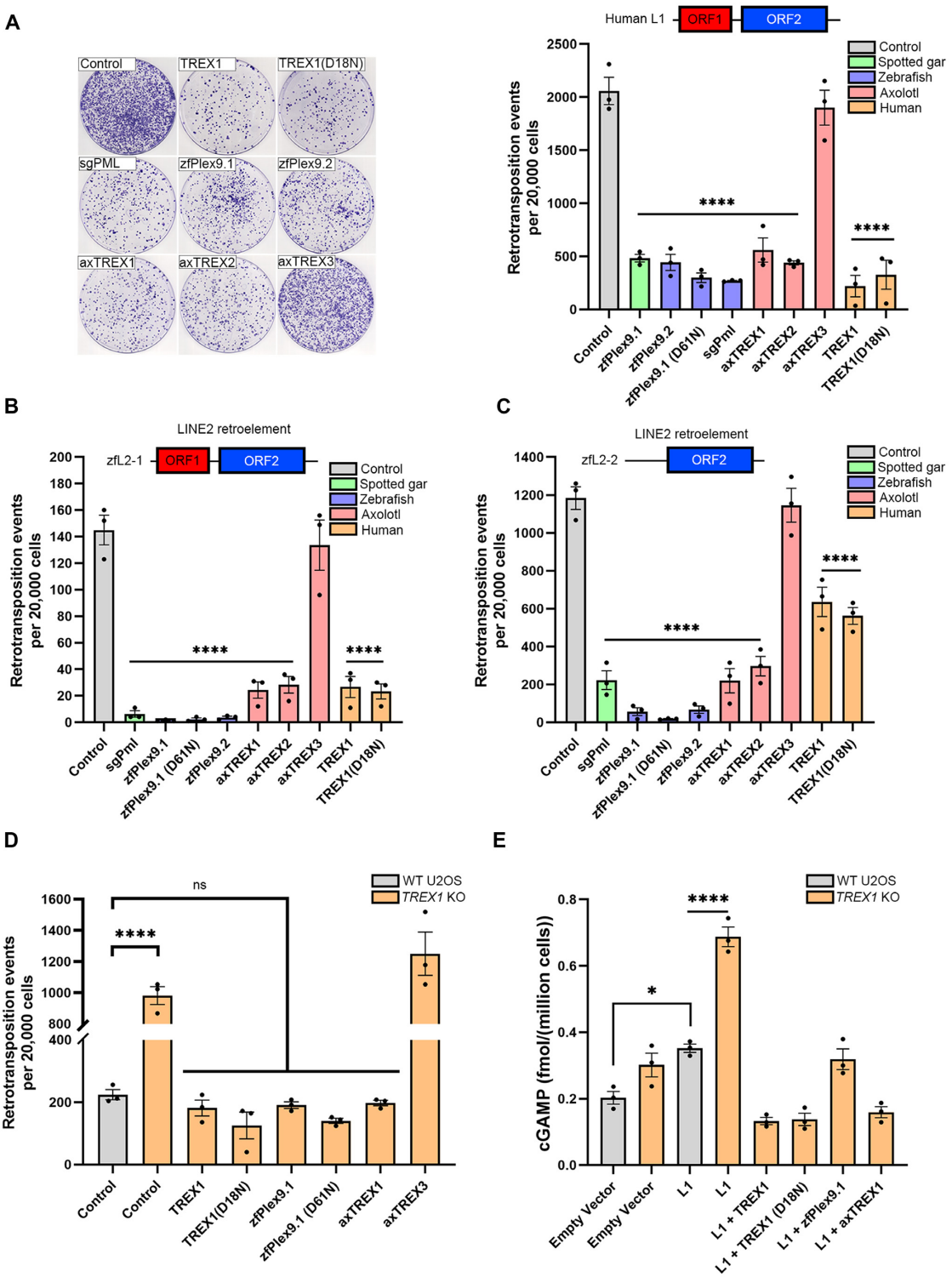

Figure 5. sgPml, Plex9 and axTREX1 proteins suppress LINE-1 (L1) activity in an exonuclease-independent manner. (**A**) Retrotransposition of human L1 was assessed with the co-expression of the FLAG-tagged versions of the indicated proteins. Resultant plates from the assay (right) were stained with 0.5% crystal violet and quantified. Retrotransposition activity of zebrafish L1 elements were also assessed in (**B**, left) zfL2-1 and (**C**, right) zfL2-2 ($n = 3$ for human L1, zfL2-1 and zfL2-2). (**D**) A *TREX1* knockout was generated using CRISPR/Cas9 and validated in U2OS (details in Supplementary Figure S12). (D) Retrotransposition activity was elevated in *TREX1* knockouts and addback of the zfPlex9.1 and zfPlex9.2, as well as axTREX1 and axTREX2 ($n = 3$). (**E**) 2′,3′-cGAMP levels in U2OS WT and *TREX1* knockout cells after the transfection of an empty vector and the human L1 vector (n = 3). 2′,3′-cGAMP concentrations were normalized to cell number for each experiment. All proteins were expressed with a FLAG epitope with the exception of GFP-TREX1(D18N). Controls represent empty vector (FLAG-tag backbone) used in the co-transfection with the L1 plasmid. All experiments were analysed by a repeated measures one-way ANOVA and then a Tukey's multiple comparisons test between groups. Asterisks indicate the following: ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

catalytic residues (Supplementary Figure S5A) that allowed us to phenocopy the exonuclease deficient D18N TREX1 mutation by mutating the cognate residue in zfPlex9.1 (i.e. D61N; Supplementary Figure S5A). The D61N mutation impaired zfPlex9.1 catalytic activity without affecting the ability to bind both ssDNA and dsDNA (Supplementary Figure S9C-F). We also attempted to disrupt zfPlex9.1 dsDNA binding by introducing an aligning TREX1 mutation known to disrupt DNA binding (R128H for TREX1; R153H for zfPlex9.1) but this did not affect the binding of zfPlex9.1 to DNA (Supplementary Figure S10) (78). While zfPlex9.1 utilizes similar catalytic residues to TREX1, it is possible that it differs mechanistically in how it binds DNA. Therefore, we found that zfPlex9.1 cleaves ssDNA and dsDNA, which likely contributes to their ability to facilitate the clearance of micronuclei.

sgPml and plex9 proteins suppress L1 retrotransposition

TREX1 is the essential brake of the cGAS-STING pathway in mammals and has been shown to prevent genome instability by degrading cytosolic DNA and micronuclei, in addition to repressing endogenous LINE-1 (L1) retrotransposons (55,56,77). The cGAS-STING pathway is highly conserved in metazoans (79). Despite lacking a *TREX1* gene homolog, teleost fishes have a functional cGAS-STING pathway (80,81), raising the possibility that another DEDDh exonuclease may exist that functionally replaces TREX1. Given that teleost *plex9.1* and *plex9.2* genes encode active DEDDh exonucleases capable of degrading ssDNA and dsDNA that may be involved in micronuclei clearance, we hypothesized that these proteins may have functionally converged with TREX1 to compensate for its absence in the teleost lineage.

The role of TREX1 in L1 suppression is independent of catalytic exonuclease function (56). Similarly, the related SAMHD1 restriction factor, also suppresses L1 in a catalytically independent manner (82). Both TREX1 and SAMHD1 appear to restrict L1 through their nucleic acid binding functions (56,82). The Plex9 proteins and sgPml both bind nucleic acids and are active exonucleases. Given their mutual exclusivity with TREX1 in the fish genomes in which they are found, we next assessed if they functionally replace TREX1 in L1 suppression. To accomplish this, we employed the neomycin-resistance based clonogenic L1 assay that was previously used to demonstrate the role of TREX1 in regulating L1 (Supplementary Figure S11A) (77), to determine if these zfPlex9 exonucleases, axTREX1/2/3, and the related sgPml, could also suppress L1 retrotransposition. As reported previously, TREX1 and exonuclease deficient TREX1 (D18N) both significantly suppressed human L1 activity significantly ($P < 0.0001$, Figure 5A). Consistent with a convergent role in the suppression of L1 retrotransposition, zfPlex9.1 and zfPlex9.2, axTREX1 and axTREX2, and sgPml all potently suppressed L1 activity (Figure 5A). Like TREX1 (D18N), exonuclease-deficient zfPlex9.1 (D61N) also suppressed L1 retrotransposition (Figure 5A).

While L1 is active in mammals, related endogenous retroelements exist in other metazoan lineages. For example, the related LINE-2 (L2) elements are active in echinoderms and teleost fishes, but not mammals. Several teleost LINE elements have been characterized (83). We wanted to determine if zfPlex9 and sgPml proteins could suppress these active L2 elements. We again utilized neomycin-based retrotransposon reporters encoding teleost L2s (zfL2-1, zfL2-2 and UnaL2) (84). Ray-finned zfPlex9.1, zfPlex9.2, and sgPml potently suppressed the retrotransposition of both zfL2-1 and zfL2-2 (Figure 5B, C). Amphibian axTREX1 and axTREX2, also suppressed the teleost L2 elements like the Plex9 proteins (Figure 5B, C). However, the truncated axTREX3 protein had no effect on L1 or L2 suppression (Figure 5A–C). We observed a similar relationship between the Plex9 proteins and axTREX1 with respect to UnaL2 (Supplementary Figure S11B). Human TREX1, although capable of suppressing zfL2-1 and zfL2-2, was not as efficient at suppressing these L2 elements compared to the teleost Plex9 proteins, suggesting a degree of species tropism between DEDDh exonucleases and their target retroelements. Therefore, it appears that Plex9 and TREX1 orthologs broadly inhibit LINE retroelements, including L2 elements found in other vertebrate species.

We next sought to determine if the Plex9 proteins could functionally complement the increase in L1 activity previously demonstrated for TREX1 loss in human cells (56,77). Thus, we generated TREX1 knockout (KO) U2OS lines (Supplementary Figure S12). As expected, TREX1 KO led to a significant increase in human L1 retrotransposition, which was reversed by the overexpression of FLAG-TREX1 or GFP-TREX1(D18N) (Figure 5D). Consistent with being true TREX1 orthologs in axolotl, the axTREX1 and axTREX2 proteins encoding full-length DEDDh domains also successfully inhibited L1 retrotransposition when over expressed in TREX1 KO U2OS cells (Figure 5D). The ectopic expression of zfPlex9.1 and zfPlex9.2 also significantly reduced human L1 retrotransposition in the TREX1 KO cell line (Figure 5D). Thus, zfPlex9.1 and zfPlex9.2 have a similar convergent function to TREX1 with respect to suppression of LINE retroelements. TREX1 inhibits the production of 2′,3′-cGAMP by cGAS after L1 activity occurs (85). We found that the significant increase in 2′,3′-cGAMP caused by TREX1 loss with the transfection of the L1 plasmid could be reversed by the addback of zfPlex9.1 and axTREX1 (Figure 5E). Together, these results are consistent with an early role for TREX1 in the suppression of L1 activity in amniotes, and that in species that lack TREX1 such as zebrafish and other ray-finned fish, Plex9 proteins converged in function to similarly suppress L1 retroelements and downstream cGAS-STING signalling.

Human PML-I shuttles to the cytoplasm to restrict L1 activity

While human PML retained the DEDDh exonuclease fold, the catalytic residues are changed relative to sgPml, and there is no evidence to suggest it is an active exonuclease (Figure 2B). To date no cellular function has been ascribed to the PML-I-CDE (Figure 6A). Since we and others have shown that L1-suppression is an exonuclease-independent function for TREX1 and the Plex9 proteins (Figure 5), we hypothesized that human PML may have retained a role in

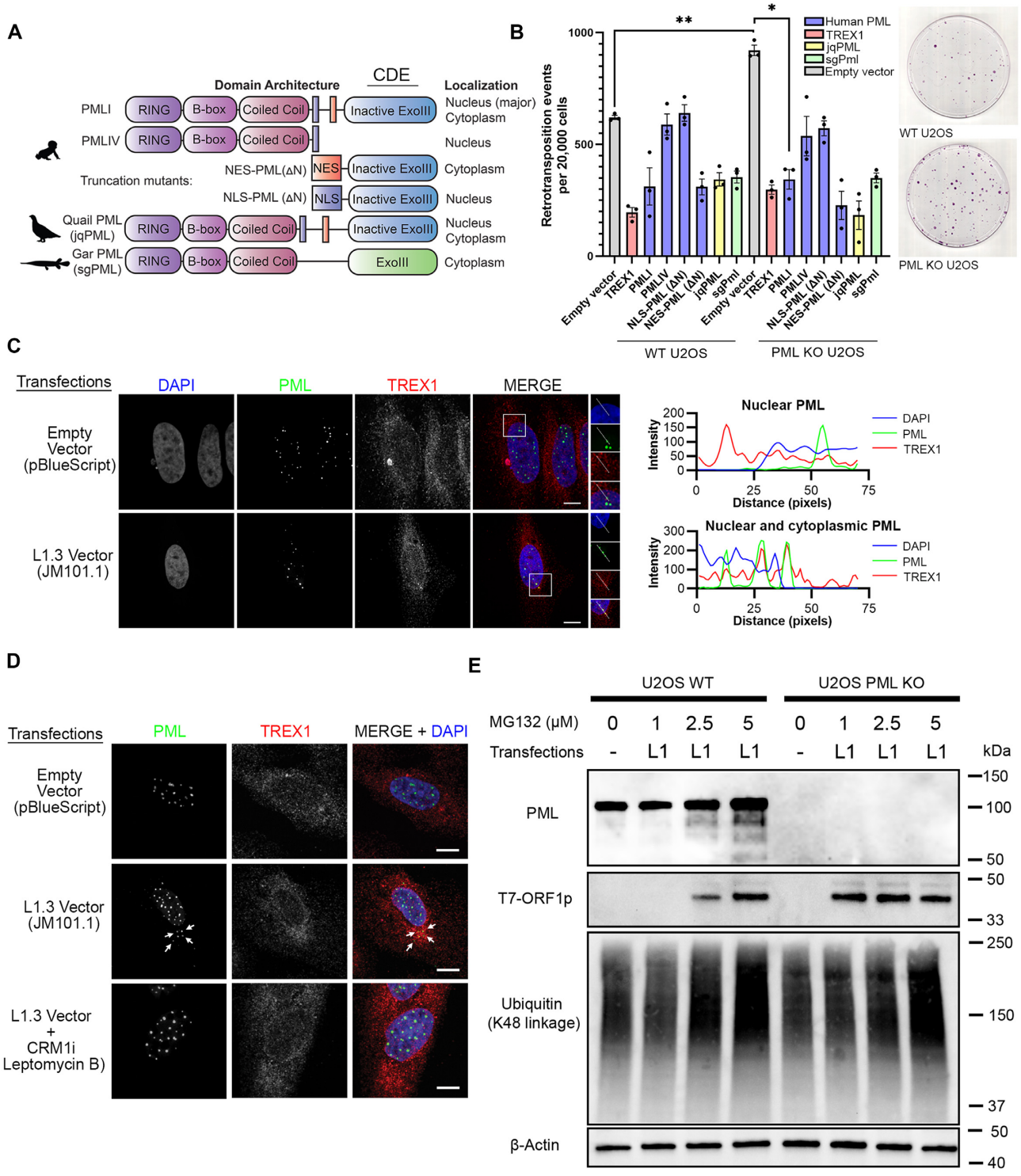

Figure 6. PML-I shuttles to the cytoplasm to suppress L1 by promoting the degradation of ORF1p. (**A**) PML mutants were generated encompassing the CDE (aa positions 596–882) and tagged with both an SV40 NLS and the PML-I NES at the N-terminus. Localizations of the mutants are shown in Supplementary Figure S14A. The mutants lack the C-terminal RBCC domain of PML. (**B**) Loss of *PML* elevates L1 activity which can be reversed with the addback of PML-I and the NES-PML mutant ($n = 3$). Resultant plates from the assay were stained with 0.5% crystal violet and quantified. Data was analysed by a repeated measures one-way ANOVA and then a Tukey's multiple comparisons test between groups. Asterisks indicate the following: $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$. (**C**) Transfection of the human L1 retrotransposition vector led to cytoplasmic PML puncta forming. PML co-localized partly with TREX1 in the cytoplasm after L1 retroelements were active. (**D**) The nucleocytoplasmic shuttling of PML-I is CRM1-dependent. U2OS cells expressing GFP-PML-I were transfected with empty vector or L1 expression vector (L1.3, JM101.1) treated for 3 h with or without CRM1 inhibitor Leptomycin B (10 ng/ml). Cytoplasmic puncta of GFP-PML-I are indicated by the white the arrows. Scale bars represent 10 μm for images in (C) and (D). (**E**) U2OS cells (WT and PML KO) were transfected with an empty vector (-) or the L1 element with T7-tagged ORF1p. Cells were then treated with different concentrations of ubiquitin-proteasome inhibitor MG132 for 10 h and T7-ORF1p protein levels were examined.

suppressing L1 activity via the CDE encoded in exon 9 of PML-I.

To test this hypothesis, we co-transfected HeLa cells with the human L1 reporter with FLAG-tagged human PML-I, PML-IV or jqPML; the latter to demonstrate conserved PML function in L1 suppression in other amniotes. PML-IV is almost identical to PML-I but lacks the CDE domain encoded by exon 9 (aa positions 596–882) (Figure 6A). We found that PML-I, but not PML-IV, significantly suppressed L1 activity ($P < 0.05$) (Supplementary Figure S13A, B). We next determined if the loss of endogenous PML influenced L1 retrotransposition in PML KO U2OS cells (13). KO of PML in U2OS cells increased L1 activity nearly 2-fold ($P < 0.01$) compared to WT cells, which was reversed by expression of PML-I, jqPML or sgPml (Figure 6B). Collectively these data indicate that PML orthologs from 3 distinct vertebrate species, quail, human and gar, are capable of robustly suppressing human L1 activity.

In addition to encoding a C-terminal DEDDh exonuclease-like domain, PML-I also uniquely encodes a nuclear export signal (NES) for which no known stimulus has been identified to trigger nucleocytoplasmic shuttling (43). We observed PML, but not PML NB component DAXX, colocalize with Clover-sgPml in the cytoplasm of a small population of stressed cells exhibiting micronuclei (Supplementary Figure S14B, C). Therefore, we next sought to examine PML nucleocytoplasmic shuttling in the context of stress induced by L1 retrotransposition. Indeed, in cells transfected with the active human L1 retroelement, endogenous PML shuttled robustly to the cytoplasm (Figure 6C). In the cytoplasm, PML formed large SUMO-negative puncta that resembled cytoplasmic sgPml bodies, which also co-localized with TREX1 puncta (Figure 6C). The expression of an inactive L1 retroelement that was not capable of retrotransposition (ORF2p D205A mutant) did not cause PML shuttling to the cytoplasm (Supplementary Figure S13C) (86). Similar to the absence of TREX1, we observed that 2′,3′-cGAMP levels were elevated significantly ($P < 0.01$) in the absence of PML after L1 transfection (Supplementary Figure S13D). This increase in 2′,3′-cGAMP levels could consequently be suppressed by the addback of PML-I expression in the PML KO cells (Supplementary Figure S13D).

Since PML-I encodes a putative NES (positions 704–713) and has been identified as an interactor of CRM1/XPO1 (66,87), we next examined if PML shuttling required CRM1, a major nuclear export protein involved in NES-dependent shuttling. To determine if PML-I shuttling was dependent on CRM1, we utilized the CRM1-inhibitor leptomycin B (88). In the presence of leptomycin B, we found that PML-I shuttling to the cytoplasm did not occur (Figure 6D). Together, these data indicate that PML nucleocytoplasmic shuttling occurs in a CRM1-dependent manner in response to active L1 retrotransposition.

Substantial PML protein is still retained in the nucleus even under conditions of high L1 activity, it raises the possibility of a nuclear function for the DEDDh domain of PML in suppressing L1 retrotransposition. To determine if this is the case, we generated N-terminus truncation mutants of PML that encoded the PML-I C-terminus (aa 597–882) with either the upstream putative NES or a nuclear localization signal from SV40 (SV40-NLS). We observed that expression of the NES-PML mutant potently suppressed L1 activity, in a manner comparable to jqPML and sgPml (Figure 6B). In contrast, the constitutively nuclear NLS-PML-I C-terminus had no significant effect on L1 retrotransposition (Figure 6B). We examined if PML localized to stress granules marked by GFP-G3BP1 and P-bodies marked by GFP-DCP1, two structures where L1 ORF1p is known to localize in the cytoplasm (61). However, cytoplasmic puncta of PML do not co-localize with either stress-granules or P-bodies after L1 transfection (Supplementary Figure S15A, B). Nonetheless, cytoplasmic PML did localize to ORF1p condensates, indicating a possible role for PML-I in the suppression of L1 ribonucleoproteins at ORF1p condensates not associated with stress granules and P-bodies (Supplementary Figure S15C).

Previous work on TREX1 revealed that it suppresses L1 by depleting ORF1p through the ubiquitin-proteasome system. Consistent with PML playing a role in ORF1p degradation, after L1 transfection we observed a higher expression of T7-tagged ORF1p in PML KO compared to PML WT U2OS cells (Figure 6E). We then assessed if this was due to impaired ORF1p degradation through the proteasome by treating L1 transfected cells with proteasome inhibitor MG132 (Figure 6E). MG132 treatment of WT cells resulted in the accumulation of T7-ORF1p as concentrations of MG132 increased, as expected. In the PML KO cells, T7-ORF1p protein levels did not change with increasing concentrations of MG132 treatment (Figure 6E). Further, addback of PML-I but not PML-IV to PML KO cells reduced the accumulation of T7-ORF1p (Supplementary Figure S13E). These data indicate that PML-I promotes the proteasome-mediated degradation of ORF1p to suppress L1. These results collectively indicate an evolutionary conserved role for PML in suppressing L1 elements, which it accomplishes by CRM1-dependent nucleocytoplasmic shuttling of PML-I to then enhance the degradation of L1 ORF1p in the cytoplasm.

DISCUSSION

This study of PML gene evolution highlights how the availability of newly sequenced genomes from diverse taxa can help illuminate the complex molecular evolution of vertebrate genes and the functions they encode. Here, we determined that PML-I encodes a vestigial DEDDh exonuclease domain that has a novel evolutionarily conserved function in L1 suppression in humans and other amniotes. The PML protein has changed significantly over the course of jawed vertebrate evolution, both in cellular localization and function, which seems to be intricately tied to its SUMOylation. For example, human PML-I is primarily a nuclear protein acting as a scaffold for PML NB formation, while gar and paddlefish PML localize to the cytoplasm (Figure 7). Nonetheless, human PML-I can still facilitate L1 suppression through the acquired ability to shuttle to the cytoplasm via its nuclear export signal. This feature represents a remarkable compensation mechanism that allows PML to retain its key role in surveillance and suppression of L1 retroelement activity despite its divergence during amniote

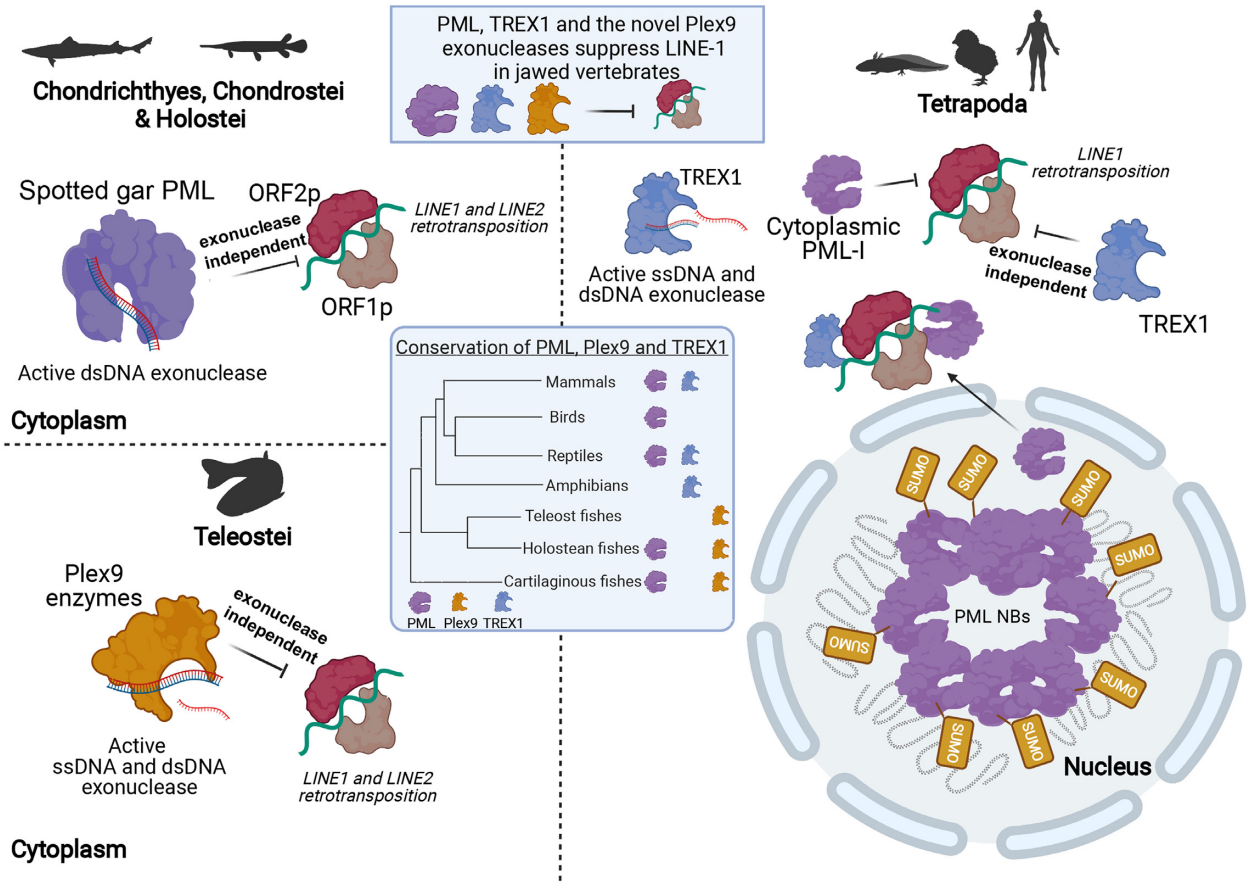

Figure 7. Overview of the suppression of LINE-1 retrotransposition in jawed vertebrates through the collective functions of PML, Plex9 and TREX1 enzymes. In fish, PML and novel Plex9 enzymes play a major role in suppressing both LINE-1 and LINE-2 elements. Despite the loss of *PML* in Teleost fish genomes, they have retained *Plex9* genes for the purpose of restricting LINE retrotransposition. *TREX1* appears later in tetrapods, where *Plex9* genes are lost. In addition, enzymatic PML function is lost and the protein forms nuclear bodies rather than localizing to the cytoplasm. However, PML has a retained an exonuclease-independent role for the surveillance of LINEs by shuttling to the cytoplasm to suppress the LINE1 retrotransposition. Thus, the three protein families (PML, Plex9 and TREX) share an important function in maintaining genome integrity in jawed vertebrates by repressing LINE-1 propagation.

evolution, which is consistent with the broader role of PML in antiviral innate immune pathways (3,4).

In studying the evolution of the PML gene, our work has implication for our understanding of how non-membrane bound organelles may have evolved in eukaryotic cells. The origins of eukaryote organelles are largely understood from the perspective of endosymbiosis. However, there are organelles that are likely derived from autogenous organellogenesis (i.e. recycling machinery and digestive vacuoles) (89). The nucleus is a complex organelle that includes a variety of subnuclear domains within, including PML NBs, Cajal bodies, paraspeckles and splicing speckles, that contribute to its architecture and function (90,91). PML NBs are membrane-less and represent molecular condensates that self-organize into organelles, in a process regulated by SUMOylation and likely occurring by phase separation (83,92,93). PML NBs emerged in amniotes coincident with the appearance of TREX1. Amniote PML neofunctionalization appears to have occurred from the loss of exonuclease function and gain of new self-organization function primed by SUMOylation (e.g. PML K616) and its localization in the nucleus. Importantly, our work highlights how membrane-less organelles not derived from endosymbiosis could originate from a single gene encoding a protein with the capacity for self-organization such as PML, which has led to the amniote innovation of the PML NB.

The cellular roles of cytoplasmic PML have been relatively understudied, particularly PML-I whose role in the cytoplasm has remained a mystery for over a decade since its first observation by Condemine and colleagues (42). We discovered that L1 retroelement transfection results in reproducible shuttling of PML-I to the cytoplasm in a manner that requires the NES of PML and is CRM1-dependent. Furthermore, ray-finned fish, turtle and avian PML orthologs which localize primarily and partially to the cytoplasm (respectively), can robustly inhibit L1 activity, indicating that this cytoplasmic function of PML was conserved over the course of amniote evolution. PML has also been reported to localize to the cytoplasm in response to HIV-1 infection and during epithelial-mesenchymal transition (EMT) (87,94). It will be of interest to determine if in these contexts PML nucleocytoplasmic shuttling is also CRM1-dependent and/or related to L1 activity.

The widely distributed DEDDh exonucleases play important roles in DNA replication, DNA repair, RNA maturation, and RNA turnover (95–98). Here, we describe and characterize the teleost Plex9 proteins as a novel subfamily of DEDDh exonucleases that have converged on a similar function in L1 suppression and micronuclei clearance to TREX1. We also found that while amniote PML orthologs do not encode a DEDDh exonuclease, that ray-finned spotted gar Pml is an active exonuclease. Importantly, we demonstrate that the Plex9 proteins are not evolutionarily related to the emergence of TREX1 orthologs in amniotes, rather they are most likely derived from a separate progenitor PML-like exonuclease. Thus, the shift in PML function from exonuclease to a scaffold-like protein underlying PML NB formation was likely a consequence of redundancy in exonuclease function with the emergence of TREX1.

Our work also has implications for our understanding of the coevolution of retrotransposons and host restriction factors that restrict them in vertebrate species. While L1s contribute to genetic variation, this is juxtaposed to their negative impact on genome integrity (22,99). As a result, several mechanisms have evolved to safeguard genome stability by preventing L1 propagation, including their suppression by TREX1 (55,77). In lobe-finned vertebrates, TREX1 appears to be a major suppressor of L1s. However, prior to the emergence of TREX1, PML and Plex9 proteins have been important contributors to safeguard genomes from L1 propagation in fish. Human PML also suppresses L1 by promoting the degradation of ORF1p. The highly conserved cGAS-STING pathway also appears to be tied to how host factors restrict or enhance L1, which has implications for innate immune signalling and senescence (25,29,77). Consistent with this aspect of L1 suppression, PML also shares the ability to inhibit L1-induced cGAS activation. Thus, DEDDh exonucleases have a unifying role in suppressing L1 retrotransposition that, surprisingly, is also independent of their exonuclease function.

While PML and TREX1 are both well-studied, their dynamic subcellular localization is less understood, and our study raises the question of why amniote PML retains the ability to suppress L1 elements? We found that TREX1 and PML co-localize in the cytoplasm when L1 elements are overexpressed and promote the degradation of L1 ORF1p. This leads us to conjecture that PML and TREX1 are mechanistically connected and contribute cooperatively to genome stability by suppressing retroelements, potentially via common SUMO-dependent and -independent protein interactions. Overall, our study has uncovered the convergent evolution of the Plex9 exonucleases in the teleost fish lineage that lacks TREX1 and PML orthologs, and a primordial role for PML in the suppression of L1 retrotransposition that is conserved across ∼350 million years of evolution.

DATA AVAILABILITY

All genomes and transcriptomes used in the study for identifying homolog sequences are available from the National Center for Biotechnology Information (NCBI). Accession numbers and identities for sequences utilized are listed in the Materials and Methods section.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Brett Racicot and Andrew W. Thompson (Michigan State University) for raising and sampling gar tissues. We also thank Dr Nika Lovšin (University of Ljubljana) and Dr Aurelien Doucet (CNRS) for providing zebrafish LINE-2 encoding plasmids and tagged LINE-1 ORF plasmids, respectively.

FUNDING

Discovery Grants from the Natural Sciences and Engineering Research Council of Canada [NSERC, RGPIN 2020-04034 to G.D., RGPIN-2017-05338 to D.L.]; G.D. is a senior scientist of the Beatrice Hunter Cancer Research Institute (BHCRI); E.H. is a trainee in the Cancer Research Training Program of the BHCRI, with funds provided by GIVETOLIVE and The Linnea Veinotte Memorial Graduate Studentship; S.M. and K.L.V. are supported by both Killam Doctoral and Dalhousie University's Presidents Awards. MP is supported by a Dalhousie Medical Research Foundation (DMRF) - Vardy Graduate Studentship; M.P. and S.M. are also supported by a Nova Scotia Graduate Scholarship; Gar work is supported by NIH [R01OD011116]; NSF EDGE program award [2029216 to I.B.]. Funding for open access charge: Discovery Grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN 2020-04034 and RGPIN-2017-05338].

Conflict of interest statement. None declared.

REFERENCES

1. de The,H., Lavau,C., Marchio,A., Chomienne,C., Degos,L. and Dejean,A. (1991) The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell*, **66**, 675–684.
2. Kakizuka,A., Miller,W.H. Jr, Umesono,K., Warrell,R.P. Jr, Frankel,S.R., Murty,V.V., Dmitrovsky,E. and Evans,R.M. (1991) Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML. *Cell*, **66**, 663–674.
3. Scherer,M. and Stamminger,T. (2016) Emerging role of PML nuclear bodies in innate immune signaling. *J. Virol.*, **90**, 5850–5854.
4. Geoffroy,M.C. and Chelbi-Alix,M.K. (2011) Role of promyelocytic leukemia protein in host antiviral defense. *J. Interferon Cytokine Res.*, **31**, 145–158.
5. Dellaire,G. and Bazett-Jones,D.P. (2004) PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *Bioessays*, **26**, 963–977.
6. Chen,M., Zhang,J., Sampieri,K., Clohessy,J.G., Mendez,L., Gonzalez-Billalabeitia,E., Liu,X.S., Lee,Y.R., Fung,J., Katon,J.M. *et al.* (2018) An aberrant SREBP-dependent lipogenic program promotes metastatic prostate cancer. *Nat. Genet.*, **50**, 206–218.
7. Maarifi,G., Chelbi-Alix,M.K. and Nisole,S. (2014) PML control of cytokine signaling. *Cytokine Growth Factor Rev.*, **25**, 551–561.
8. Bernardi,R. and Pandolfi,P.P. (2007) Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.*, **8**, 1006–1016.

9. Bernardi,R., Papa,A., Egia,A., Coltella,N., Teruya-Feldstein,J., Signoretti,S. and Pandolfi,P.P. (2011) Pml represses tumour progression through inhibition of mTOR. *EMBO Mol. Med.*, **3**, 249–257.
10. Killcoyne,S., Yusuf,A. and Fitzgerald,R.C. (2021) Genomic instability signals offer diagnostic possibility in early cancer detection. *Trends Genet.*, **37**, 966–972.
11. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
12. Chang,H.R., Munkhjargal,A., Kim,M.J., Park,S.Y., Jung,E., Ryu,J.H., Yang,Y., Lim,J.S. and Kim,Y. (2018) The functional roles of PML nuclear bodies in genome maintenance. *Mutat. Res.*, **809**, 99–107.
13. Attwood,K.M., Salsman,J., Chung,D., Mathavarajah,S., Van Iderstine,C. and Dellaire,G. (2020) PML isoform expression and DNA break location relative to PML nuclear bodies impacts the efficiency of homologous recombination. *Biochem. Cell. Biol.*, **98**, 314–326.
14. Ugge,M., Simoni,M., Fracassi,C. and Bernardi,R. (2022) PML isoforms: a molecular basis for PML pleiotropic functions. *Trends Biochem. Sci.*, **47**, 609–619.
15. Chang,F.T., McGhie,J.D., Chan,F.L., Tang,M.C., Anderson,M.A., Mann,J.R., Andy Choo,K.H. and Wong,L.H. (2013) PML bodies provide an important platform for the maintenance of telomeric chromatin integrity in embryonic stem cells. *Nucleic Acids Res.*, **41**, 4447–4458.
16. Cordaux,R. and Batzer,M.A. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.*, **10**, 691–703.
17. Han,K., Sen,S.K., Wang,J., Callinan,P.A., Lee,J., Cordaux,R., Liang,P. and Batzer,M.A. (2005) Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.*, **33**, 4040–4052.
18. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
19. Brouha,B., Schustak,J., Badge,R.M., Lutz-Prigge,S., Farley,A.H., Moran,J.V. and Kazazian,H.H. Jr (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 5280–5285.
20. Beck,C.R., Collier,P., Macfarlane,C., Malig,M., Kidd,J.M., Eichler,E.E., Badge,R.M. and Moran,J.V. (2010) LINE-1 retrotransposition activity in human genomes. *Cell*, **141**, 1159–1170.
21. Belancio,V.P., Roy-Engel,A.M., Pochampally,R.R. and Deininger,P. (2010) Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.*, **38**, 3909–3922.
22. Gasior,S.L., Wakeman,T.P., Xu,B. and Deininger,P.L. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.*, **357**, 1383–1393.
23. Goodier,J.L. and Kazazian,H.H. Jr (2008) Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, **135**, 23–35.
24. Muotri,A.R., Chu,V.T., Marchetto,M.C., Deng,W., Moran,J.V. and Gage,F.H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, **435**, 903–910.
25. Liang,C., Ke,Q., Liu,Z., Ren,J., Zhang,W., Hu,J., Wang,Z., Chen,H., Xia,K., Lai,X. *et al.* (2022) BMAL1 moonlighting as a gatekeeper for LINE1 repression and cellular senescence in primates. *Nucleic Acids Res.*, **50**, 3323–3347.
26. Horn,A.V., Klawitter,S., Held,U., Berger,A., Vasudevan,A.A., Bock,A., Hofmann,H., Hanschmann,K.M., Trosemeier,J.H., Flory,E. *et al.* (2014) Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. *Nucleic Acids Res.*, **42**, 396–416.
27. Brooks,W. (2020) An epigenetics-based hypothesis of autoantigen development in systemic lupus erythematosus. *Epigenomes*, **4**, 6.
28. Gamdzyk,M., Doycheva,D.M., Araujo,C., Ocak,U., Luo,Y., Tang,J. and Zhang,J.H. (2020) cGAS/STING pathway activation contributes to delayed neurodegeneration in neonatal hypoxia-ischemia rat model: possible involvement of LINE-1. *Mol. Neurobiol.*, **57**, 2600–2619.
29. Bregnard,C., Guerra,J., Dejardin,S., Passalacqua,F., Benkirane,M. and Laguette,N. (2016) Upregulated LINE-1 activity in the Fanconi anemia cancer susceptibility syndrome leads to spontaneous pro-inflammatory cytokine production. *EBioMedicine*, **8**, 184–194.
30. Lim,Y.W., Sanz,L.A., Xu,X., Hartono,S.R. and Chedin,F. (2015) Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. *Elife*, **4**, e08007.
31. Terry,D.M. and Devine,S.E. (2019) Aberrantly high levels of somatic LINE-1 expression and retrotransposition in human neurological disorders. *Front. Genet.*, **10**, 1244.
32. Benitez-Guijarro,M., Lopez-Ruiz,C., Tarnauskaite,Z., Murina,O., Mian Mohammad,M., Williams,T.C., Fluteau,A., Sanchez,L., Vilar-Astasio,R., Garcia-Canadas,M. *et al.* (2018) RNase H2, mutated in Aicardi-Goutieres syndrome, promotes LINE-1 retrotransposition. *EMBO J.*, **37**, e98506.
33. Zhao,K., Du,J., Han,X., Goodier,J.L., Li,P., Zhou,X., Wei,W., Evans,S.L., Li,L., Zhang,W. *et al.* (2013) Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep.*, **4**, 1108–1115.
34. Volkman,H.E. and Stetson,D.B. (2014) The enemy within: endogenous retroelements and autoimmune disease. *Nat. Immunol.*, **15**, 415–422.
35. Choi,J., Hwang,S.Y. and Ahn,K. (2018) Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition. *Nucleic Acids Res.*, **46**, 1912–1926.
36. Goodier,J.L., Cheung,L.E. and Kazazian,H.H. Jr (2012) MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLos Genet.*, **8**, e1002941.
37. Orecchini,E., Doria,M., Antonioni,A., Galardi,S., Ciafre,S.A., Frassinelli,L., Mancone,C., Montaldo,C., Tripodi,M. and Michienzi,A. (2017) ADAR1 restricts LINE-1 retrotransposition. *Nucleic Acids Res.*, **45**, 155–168.
38. Crow,Y.J., Hayward,B.E., Parmar,R., Robins,P., Leitch,A., Ali,M., Black,D.N., van Bokhoven,H., Brunner,H.G., Hamel,B.C. *et al.* (2006) Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat. Genet.*, **38**, 917–920.
39. Lovsin,N., Gubensek,F. and Kordi,D. (2001) Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol. Biol. Evol.*, **18**, 2213–2224.
40. Volff,J.N. (2005) Genome evolution and biodiversity in teleost fish. *Heredity (Edinb)*, **94**, 280–294.
41. Veinotte,C.J., Dellaire,G. and Berman,J.N. (2014) Hooking the big one: the potential of zebrafish xenotransplantation to reform cancer drug screening in the genomic era. *Dis. Model. Mech.*, **7**, 745–754.
42. Condemine,W., Takahashi,Y., Le Bras,M. and de The,H. (2007) A nucleolar targeting signal in PML-I addresses PML to nucleolar caps in stressed or senescent cells. *J. Cell Sci.*, **120**, 3219–3227.
43. Condemine,W., Takahashi,Y., Zhu,J., Puvion-Dutilleul,F., Guegan,S., Janin,A. and de The,H. (2006) Characterization of endogenous human promyelocytic leukemia isoforms. *Cancer Res.*, **66**, 6192–6198.
44. Garcia-Perez,J.L., Widmann,T.J. and Adams,I.R. (2016) The impact of transposable elements on mammalian development. *Development*, **143**, 4101–4114.
45. Grundy,E.E., Diab,N. and Chiappinelli,K.B. (2022) Transposable element regulation and expression in cancer. *FEBS J.*, **289**, 1160–1179.
46. Katoh,K., Misawa,K., Kuma,K. and Miyata,T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
47. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
48. Nguyen,L.T., Schmidt,H.A., von Haeseler,A. and Minh,B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, **32**, 268–274.
49. Darnet,S., Dragalzew,A.C., Amaral,D.B., Sousa,J.F., Thompson,A.W., Cass,A.N., Lorena,J., Pires,E.S., Costa,C.M., Sousa,M.P. *et al.* (2019) Deep evolutionary origin of limb and fin regeneration. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 15106–15115.
50. Pinder,J., Salsman,J. and Dellaire,G. (2015) Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.*, **43**, 9379–9392.

51. Deschamps,T. and Kalamvoki,M. (2017) Impaired STING pathway in human osteosarcoma U2OS cells contributes to the growth of ICP0-null mutant herpes simplex virus. *J. Virol.*, **91**, e00006-17,
52. Labun,K., Montague,T.G., Krause,M., Torres Cleuren,Y.N., Tjeldnes,H. and Valen,E. (2019) CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.*, **47**, W171–W174.
53. Mohr,L., Toufektchan,E., von Morgen,P., Chu,K., Kapoor,A. and Maciejowski,J. (2021) ER-directed TREX1 limits cGAS activation at micronuclei. *Mol. Cell*, **81**, 724–738.
54. Kucej,M., Fermaintt,C.S., Yang,K., Irizarry-Caro,R.A. and Yan,N. (2017) Mitotic phosphorylation of TREX1 C terminus disrupts TREX1 regulation of the oligosaccharyltransferase complex. *Cell Rep.*, **18**, 2600–2607.
55. Thomas,C.A., Tejwani,L., Trujillo,C.A., Negraes,P.D., Herai,R.H., Mesci,P., Macia,A., Crow,Y.J. and Muotri,A.R. (2017) Modeling of TREX1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. *Cell Stem Cell*, **21**, 319–331.
56. Li,P., Du,J., Goodier,J.L., Hou,J., Kang,J., Kazazian,H.H. Jr, Zhao,K. and Yu,X.F. (2017) Aicardi-Goutieres syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. *Nucleic Acids Res.*, **45**, 4619–4631.
57. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H. Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
58. Bogerd,H.P., Wiegand,H.L., Hulme,A.E., Garcia-Perez,J.L., O'Shea,K.S., Moran,J.V. and Cullen,B.R. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8780–8785.
59. Muckenfuss,H., Hamdorf,M., Held,U., Perkovic,M., Lower,J., Cichutek,K., Flory,E., Schumann,G.G. and Munk,C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.*, **281**, 22161–22172.
60. Bogerd,H.P., Wiegand,H.L., Doehle,B.P., Lueders,K.K. and Cullen,B.R. (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.*, **34**, 89–95.
61. Doucet,A.J., Hulme,A.E., Sahinovic,E., Kulpa,D.A., Moldovan,J.B., Kopera,H.C., Athanikar,J.N., Hasnaoui,M., Bucheton,A., Moran,J.V. *et al.* (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.*, **6**, e1001150.
62. Haller,M., Hock,A.K., Giampazolias,E., Oberst,A., Green,D.R., Debnath,J., Ryan,K.M., Vousden,K.H. and Tait,S.W. (2014) Ubiquitination and proteasomal degradation of ATG12 regulates its proapoptotic activity. *Autophagy*, **10**, 2269–2278.
63. He,X., Riceberg,J., Soucy,T., Koenig,E., Minissale,J., Gallery,M., Bernard,H., Yang,X., Liao,H., Rabino,C. *et al.* (2017) Probing the roles of SUMOylation in cancer cell biology by using a selective SAE inhibitor. *Nat. Chem. Biol.*, **13**, 1164–1171.
64. Brazeau,M.D. and Friedman,M. (2015) The origin and early phylogenetic history of jawed vertebrates. *Nature*, **520**, 490–497.
65. Meroni,G. (2012) Genomics and evolution of the TRIM gene family. *Adv. Exp. Med. Biol.*, **770**, 1–9.
66. Nisole,S., Maroui,M.A., Mascle,X.H., Aubry,M. and Chelbi-Alix,M.K. (2013) Differential roles of PML isoforms. *Front. Oncol.*, **3**, 125.
67. Braasch,I., Gehrke,A.R., Smith,J.J., Kawasaki,K., Manousaki,T., Pasquier,J., Amores,A., Desvignes,T., Batzel,P., Catchen,J. *et al.* (2016) The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.*, **48**, 427–437.
68. Huang,K.W., Liu,T.C., Liang,R.Y., Chu,L.Y., Cheng,H.L., Chu,J.W. and Hsiao,Y.Y. (2018) Structural basis for overhang excision and terminal unwinding of DNA duplexes by TREX1. *PLoS Biol.*, **16**, e2005653.
69. Cuchet-Lourenco,D., Boutell,C., Lukashchuk,V., Grant,K., Sykes,A., Murray,J., Orr,A. and Everett,R.D. (2011) SUMO pathway dependent recruitment of cellular repressors to herpes simplex virus type 1 genomes. *PLoS Pathog.*, **7**, e1002123.
70. Davesne,D., Friedman,M., Schmitt,A.D., Fernandez,V., Carnevale,G., Ahlberg,P.E., Sanchez,S. and Benson,R.B.J. (2021) Fossilized cell structures identify an ancient origin for the teleost whole-genome duplication. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2101780118.
71. Hemphill,W.O. and Perrino,F.W. (2019) Measuring TREX1 and TREX2 exonuclease activities. *Methods Enzymol.*, **625**, 109–133.
72. Lindahl,T., Barnes,D.E., Yang,Y.G. and Robins,P. (2009) Biochemical properties of mammalian TREX1 and its association with DNA replication and inherited inflammatory disease. *Biochem. Soc. Trans.*, **37**, 535–538.
73. Miyazaki,T., Kim,Y.S., Yoon,J., Wang,H., Suzuki,T. and Morse,H.C 3rd (2014) The 3'-5' DNA exonuclease TREX1 directly interacts with poly(ADP-ribose) polymerase-1 (PARP1) during the DNA damage response. *J. Biol. Chem.*, **289**, 32548–32558.
74. Kim,S.H., Kim,G.H., Kemp,M.G. and Choi,J.H. (2022) TREX1 degrades the 3' end of the small DNA oligonucleotide products of nucleotide excision repair in human cells. *Nucleic Acids Res.*, **50**, 3974–3984.
75. Christmann,M., Tomicic,M.T., Aasland,D., Berdelle,N. and Kaina,B. (2010) Three prime exonuclease I (TREX1) is Fos/AP-1 regulated by genotoxic stress and protects against ultraviolet light and benzo(a)pyrene-induced DNA damage. *Nucleic Acids Res.*, **38**, 6418–6432.
76. Wei,X., Wang,Z., Hinson,C. and Yang,K. (2022) Human TDP1, APE1 and TREX1 repair 3'-DNA-peptide/protein cross-links arising from abasic sites in vitro. *Nucleic Acids Res.*, **50**, 3638–3657.
77. Stetson,D.B., Ko,J.S., Heidmann,T. and Medzhitov,R. (2008) Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell*, **134**, 587–598.
78. Zhou,W., Richmond-Buccola,D., Wang,Q. and Kranzusch,P.J. (2022) Structural basis of human TREX1 DNA degradation and autoimmune disease. *Nat. Commun.*, **13**, 4277.
79. Wu,X., Wu,F.H., Wang,X., Wang,L., Siedow,J.N., Zhang,W. and Pei,Z.M. (2014) Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.*, **42**, 8243–8257.
80. Liu,Z.F., Ji,J.F., Jiang,X.F., Shao,T., Fan,D.D., Jiang,X.H., Lin,A.F., Xiang,L.X. and Shao,J.Z. (2020) Characterization of cGAS homologs in innate and adaptive mucosal immunities in zebrafish gives evolutionary insights into cGAS-STING pathway. *FASEB J.*, **34**, 7786–7809.
81. de Oliveira Mann,C.C., Orzalli,M.H., King,D.S., Kagan,J.C., Lee,A.S.Y. and Kranzusch,P.J. (2019) Modular architecture of the STING C-terminal tail allows interferon and NF-kappaB signaling adaptation. *Cell Rep.*, **27**, 1165–1175.
82. Seamon,K.J., Sun,Z., Shlyakhtenko,L.S., Lyubchenko,Y.L. and Stivers,J.T. (2015) SAMHD1 is a single-stranded nucleic acid binding protein with no active site-associated nuclease activity. *Nucleic Acids Res.*, **43**, 6486–6499.
83. Shen,T.H., Lin,H.K., Scaglioni,P.P., Yung,T.M. and Pandolfi,P.P. (2006) The mechanisms of PML-nuclear body formation. *Mol. Cell*, **24**, 331–339.
84. Lindic,N., Budic,M., Petan,T., Knisbacher,B.A., Levanon,E.Y. and Lovsin,N. (2013) Differential inhibition of LINE1 and LINE2 retrotransposition by vertebrate AID/APOBEC proteins. *Retrovirology*, **10**, 156.
85. Gao,D., Li,T., Li,X.D., Chen,X., Li,Q.Z., Wight-Carter,M. and Chen,Z.J. (2015) Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5699–E5705.
86. Farkash,E.A., Kao,G.D., Horman,S.R. and Prak,E.T. (2006) Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res.*, **34**, 1196–1204.
87. Buczek,M.E., Miles,A.K., Green,W., Johnson,C., Boocock,D.J., Pockley,A.G., Rees,R.C., Hulman,G., van Schalkwyk,G., Parkinson,R. *et al.* (2016) Cytoplasmic PML promotes TGF-beta-associated epithelial-mesenchymal transition and invasion in prostate cancer. *Oncogene*, **35**, 3465–3475.
88. Kudo,N., Matsumori,N., Taoka,H., Fujiwara,D., Schreiner,E.P., Wolff,B., Yoshida,M. and Horinouchi,S. (1999) Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9112–9117.
89. Dacks,J.B., Poon,P.P. and Field,M.C. (2008) Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 588–593.

90. Handwerger,K.E. and Gall,J.G. (2006) Subnuclear organelles: new insights into form and function. *Trends Cell Biol.*, **16**, 19–26.
91. Liu,S., Zhang,L., Quan,H., Tian,H., Meng,L., Yang,L., Feng,H. and Gao,Y.Q. (2018) From 1D sequence to 3D chromatin dynamics and cellular functions: a phase separation perspective. *Nucleic Acids Res.*, **46**, 9367–9383.
92. Banani,S.F., Rice,A.M., Peeples,W.B., Lin,Y., Jain,S., Parker,R. and Rosen,M.K. (2016) Compositional control of phase-separated cellular bodies. *Cell*, **166**, 651–663.
93. Corpet,A., Kleijwegt,C., Roubille,S., Juillard,F., Jacquet,K., Texier,P. and Lomonte,P. (2020) PML nuclear bodies and chromatin dynamics: catch me if you can!*Nucleic Acids Res.*, **48**, 11890–11912.
94. Turelli,P., Doucas,V., Craig,E., Mangeat,B., Klages,N., Evans,R., Kalpana,G. and Trono,D. (2001) Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol. Cell*, **7**, 1245–1254.
95. Zuo,Y. and Deutscher,M.P. (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.*, **29**, 1017–1026.
96. Yang,W. (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1–93.
97. Cheng,H.L., Lin,C.T., Huang,K.W., Wang,S., Lin,Y.T., Toh,S.I. and Hsiao,Y.Y. (2018) Structural insights into the duplex DNA processing of TREX2. *Nucleic Acids Res.*, **46**, 12166–12176.
98. Hsiao,Y.Y., Duh,Y., Chen,Y.P., Wang,Y.T. and Yuan,H.S. (2012) How an exonuclease decides where to stop in trimming of nucleic acids: crystal structures of RNase T-product complexes. *Nucleic Acids Res.*, **40**, 8144–8154.
99. Goodier,J.L. (2016) Restricting retrotransposons: a review. *Mob. DNA*, **7**, 16.