

# When Genomics Is Not Enough: Experimental Evidence for a Decrease in LINE-1 Activity During the Evolution of Australian Marsupials

Susanne Gallus<sup>1,2</sup>, Fritjof Lammers<sup>1,2</sup>, and Maria Anna Nilsson<sup>1</sup>

<sup>1</sup>Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft fuer Naturforschung, Senckenberg Anlage 25, Frankfurt, Germany

<sup>2</sup>Institute for Ecology, Evolution and Diversity, Faculty of Biological Sciences, Johann Wolfgang Goethe University Frankfurt Am Main, Max-von-Laue Straße 9, 60438 Frankfurt am Main, Germany

\*Corresponding author: E-mail: maria.nilsson-janke@senckenberg.de.

Accepted: June 22, 2016

**Data deposition:** Sequences were submitted via EMBL/ENA. Experimentally generated sequences have been deposited at GenBank under the accession LT593140–LT593738.

## Abstract

The autonomous transposable element LINE-1 is a highly abundant element that makes up between 15% and 20% of therian mammal genomes. Since their origin before the divergence of marsupials and placental mammals, LINE-1 elements have contributed actively to the genome landscape. A previous *in silico* screen of the Tasmanian devil genome revealed a lack of functional coding LINE-1 sequences. In this study we present the results of an *in vitro* analysis from a partial LINE-1 reverse transcriptase coding sequence in five marsupial species. Our experimental screen supports the *in silico* findings of the genome-wide degradation of LINE-1 sequences in the Tasmanian devil, and identifies a high frequency of degraded LINE-1 sequences in other Australian marsupials. The comparison between the experimentally obtained LINE-1 sequences and reference genome assemblies suggests that conclusions from *in silico* analyses of retrotransposition activity can be influenced by incomplete genome assemblies from short reads.

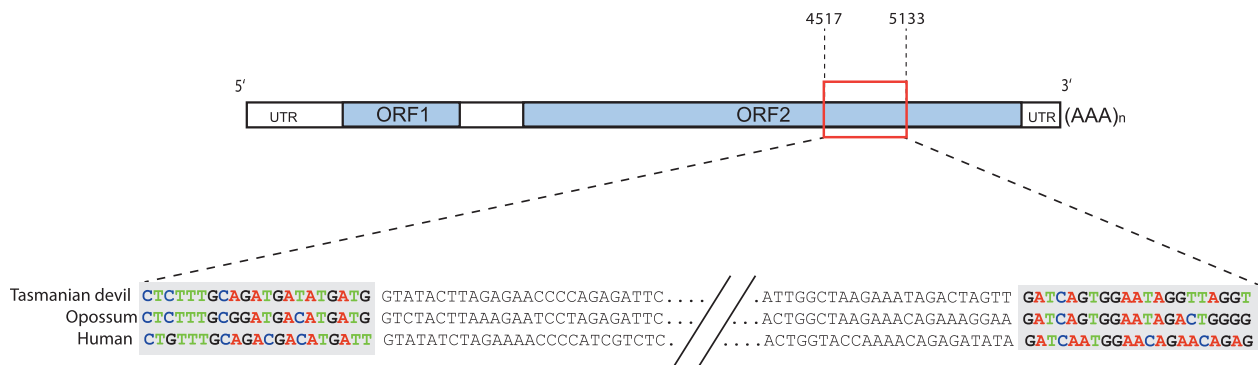
**Key words:** LINE-1, marsupials, transposable element, retrotransposition, Tasmanian devil.

## Introduction

Transposable elements (TEs) are highly abundant, varied, and play a significant role in genome evolution and structure (Oliver and Greene 2009; Chalopin et al. 2015). When analyzing high throughput sequencing data, the question is: How much information can be drawn from *in silico* analyses of current genome assemblies on the function of TEs and their interactions? It is well known—yet rarely addressed—that repetitive regions, like TEs, cause significant problems for genome assembly algorithms, which can lead to incompletely sequenced genomes (Alkan et al. 2011; Treangen and Salzberg 2012). With up to 10,000 nucleotides (nt), and between one and four open reading frames (ORFs) autonomous TEs are considerably longer than most reads from high throughput sequencing (current Illumina read length 125–250 bp) (Feng et al. 1996; Malik et al. 1999), and particularly susceptible to assembly artifacts (Treangen and Salzberg

2012). Autonomous TEs encode the protein machinery that is necessary to actively propagate TEs to new genomic locations. The retrotransposition or transposition activity of autonomous TEs is defined by the coding ORFs. Currently, no experimental protocol exists to extract unknown functional TEs from a genome, and studies are often solely based on *in silico* analyses of genome assemblies.

The autonomous non-LTR retrotransposon Long Interspersed Element-1 (LINE-1) (fig. 1) provides the enzymatic retrotransposition machinery in most mammalian genomes, which is also hitchhiked by the nonautonomous Short Interspersed Elements (SINEs) (Okada et al. 1997). It was assumed that retrotranspositionally active LINE-1 sequences would be found ubiquitously across all mammalian genomes (Shedlock et al. 2004). The maintenance of retrotranspositionally active LINE-1 copies in mammalian genomes remains interesting, because TE insertions can be deleterious, causing



**Fig. 1.**—Schematic drawing of a functional mammalian LINE-1, which is structured into five domains: a 5′ untranslated region (UTR), followed by two ORFs, a 3′-UTR, and a poly(A)tail ((AAA)<sub>n</sub>). The essential enzymes for retrotransposition are encoded on the 3′ proximal ORF (ORF2), reverse transcriptase and endonuclease (Feng et al. 1996; Malik et al. 1999). The red box indicates the 612 nt region that was PCR amplified and analyzed. The numbers above the red box is the relative nt position of the region in the Tasmanian devil LINE-1 sequence. Species-specific primer sequences for PCR amplification are indicated in the grey shaded boxes.

various diseases, or be beneficial in adaptive evolution (Mills et al. 2007; González and Petrov 2009). Results from an in silico TE screen of a marsupial carnivore genome, the Tasmanian devil (*Sarcophilus harrisii*) suggested that, although LINE-1 activity had been silenced (Gallus et al. 2015). Thus, the findings from the Tasmanian devil genome analysis challenged the notion of a general preservation of LINE-1 activity in all mammals. The possibility remains, however, that the apparent lack of full-length LINE-1 copies, of approximately 6,000 nt, may be the result of an incorrect assembly of the two ORFs (Gallus et al. 2015; Nilsson 2016).

The South American sigmodontine rodents were reported as the first mammalian taxon to exhibit LINE-1 silencing (Casavant et al. 2000; Grahn et al. 2005; Rinehart et al. 2005). The second discovery of a loss of LINE-1 activity about 24 million years ago (Ma) was reported in the species-rich megabats (Cantrell et al. 2008). A LINE-1 inactivation event was reported for the 13-lined ground squirrel, and finally strongly reduced LINE-1 activity was suggested for the spider monkey (Boissinot et al. 2004; Platt and Ray 2012).

Here we apply an experimental protocol (Cantrell et al. 2000) to analyze the fraction of putatively functional to non-functional LINE-1 copies in the Tasmanian devil genome and other dasyuromorphian marsupials. As a control, we performed the same LINE-1 screen in the opossum and human genome, both of which are known to contain functional LINE-1 sequences (Gu et al. 2007; Stewart et al. 2011). Comparative genomics in mammals has identified a highly conserved 612 nt region in the ORF2 (~3,800 nt) of functional LINE-1 elements, where the correct reading frame, without stop codons, is indicative of LINE-1 retrotransposition capability (Cantrell et al. 2000; Scott et al. 2006) (fig. 1). Species-specific PCR primers were constructed to target the highly conserved region and preferentially amplify recent LINE-1 insertions (Cantrell et al. 2000; Grahn et al. 2005). The

PCR products were cloned, Sanger sequenced, and analyzed for frame shifts and sequence similarity for the studied species (supplementary material, Supplementary Material online). Each amplified DNA sequence was translated into its amino acid (aa) sequence to determine if they originated from potentially active LINE-1s. Sequences that contained insertions/deletions (indels) leading to frame-shifts or missense mutations causing premature stop codons indicate a nonfunctional LINE-1 copy, while sequences that were translated without stop-codons were assumed putatively retrotranspositionally active. This approach will likely overestimate the number of functional LINE-1 elements, because stop-codons might be located outside the analyzed sequence. In total we generated and analyzed 700 LINE-1 sequences from six mammalian species (table S1, supplementary material, Supplementary Material online).

### Limited Functional LINE-1 Copies in the Genome of Carnivorous Marsupials

We investigated the LINE-1 content in three dasyuromorphian species: Tasmanian devil, quoll (*Dasyurus geoffroii*), and fat-tailed dunnart (*Sminthopsis crassicaudata*). The dunnart belongs to the sub-family Sminthopsini that split from the Dasyurini (quoll/Tasmanian devil) about 30 Ma (Meredith et al. 2011). At least 100 LINE-1 sequences were generated for each of the species to minimize experimental bias (table 1). As the main focus was on the Tasmanian devil, 200 sequences were generated for this species. Only 19 sequences (9.5%) from the 200 Tasmanian devil sequences were possible to translate without premature stop codons or disrupted reading frames (table 1). Each of the 19 translatable sequences was matched to the respective genomic location in the Tasmanian devil genome assembly to study the properties of the flanking sequences. However, it was not possible to find any full-length functional LINE-1 sequences from these 19 sequences in the

**Table 1**

Results from the PCR Screen of Six Mammals

Common name	Tasmanian devil	Northern quoll	Fat-tailed dunnart	Bandicoot	Opossum	Human
Scientific name	<i>S. harrisii</i>	<i>D. geoffroii</i>	<i>S. crassicaudata</i>	<i>P. gunnii</i>	<i>M. domestica</i>	<i>Homo sapiens</i>
Number of clones	200	100	100	100	100	100
Intact ORF2 fragments	19	14	2	8	35	40
Mean group distance <sup>a</sup>	0.142	0.154	0.148	0.243	0.05	0.05

<sup>a</sup>Overall mean distance was calculated in MEGA7 and shows the number of base substitutions per site from averaging over all sequence pairs. Analyses were conducted using the Tamura–Nei model (Tamura and Nei 1993) and involved all clones for each species. Positions with less than 95% coverage were excluded.

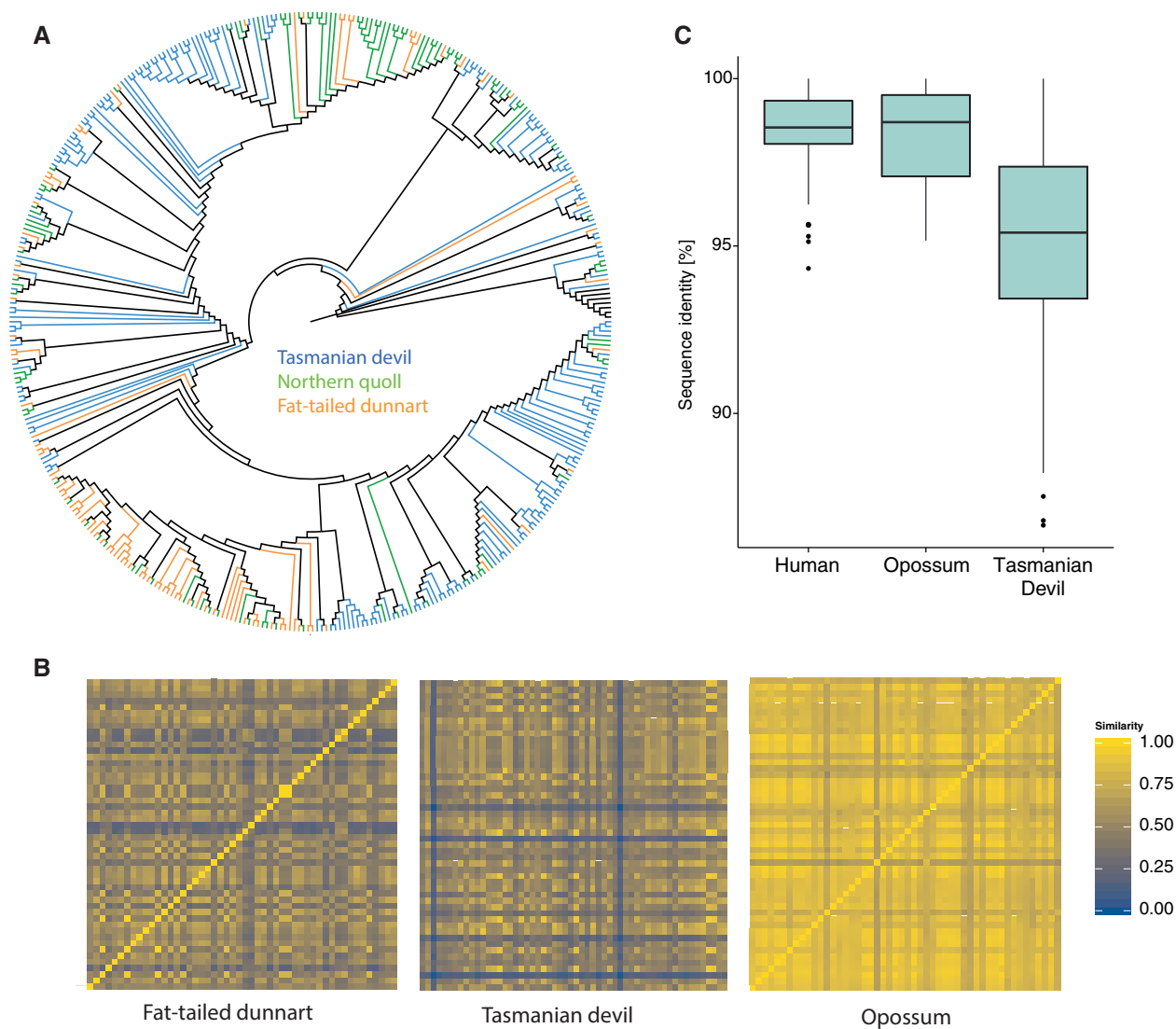
genome assembly (supplementary table S1, Supplementary Material online). Scaffolds with the most similar sequences, usually ended in the middle of the LINE-1 element, or the LINE-1 was interrupted by stretches of unresolved nucleotides (“N”).

The PCR screen of 100 clones yielded 14 clones for the quoll, and 2 clones for the fat-tailed dunnart (*S. crassicaudata*) that were possible to be translated without stop codons. This is a similar fraction of degraded LINE-1 sequences compared to the Tasmanian devil sequences. To validate the accuracy, the two PCR primers were tested in silico on published striped-faced dunnart (*Sminthopsis macroura*) sequence data. We mined public databases for LINE-1 sequences from the dunnart using nucleotide and protein queries for LINE-1 and both ORFs. From 1,085 Mb sequence, 76 mostly partial ORF2 sequences were identified, of which only four covered both primer binding sites (supplementary table S3, Supplementary Material online). However, all 76 extracted ORF2 sequences from dunnart contained stop codons, in agreement with the experimentally generated sequences. To identify species-specific LINE-1 subfamily clusters the 400 dasyuromorphian sequences were combined into a multi-species alignment. The phylogenetic maximum likelihood (ML) tree (fig. 2A) shows at least three clusters with sequences from mostly quoll, Tasmanian devil, or dunnart. This is expected because of the long evolutionary divergence between the species, and the fact that there has been LINE-1 activity in the Tasmanian devil genome since its divergence from the quoll (Gallus et al. 2015). The fact that we also found species-specific clusters for quoll and dunnart suggests that LINE-1 remained active after the lineages split. The results suggest that there was reduced LINE-1 activity in the common ancestor of Dasyuromorphia, but a final inactivation occurred independently in each species. The inactivation pattern from the phylogenetic tree of marsupials differs from that of the megabat LINE-1, where no species-specific clusters could be recovered (Cantrell et al. 2008). Our analysis of functional LINE-1s in the genomes of three dasyuromorphian species suggests that for unknown reasons the dasyuromorphian genomes are depleted of functional LINE-1 copies. It is possible that some functional LINE-1 copies remain, but in the absence of a well-assembled genome their identification will remain elusive.

### Retrotranspositionally Active Species Have Different LINE-1 Diversity than Australian Marsupials

The functional LINE-1 content in the Australian dasyuromorphian marsupial genomes appears to be different from the South American opossum (*Monodelphis domestica*). Recent experimental screening of SINE insertions in several opossum populations indicated ongoing LINE-1 retrotransposition activity (Gu et al. 2007). We designed primers specific for opossum LINE-1 sequences (Gentles et al. 2007). From 100 randomly selected and sequenced clones, 35 could be translated (table 1). This is three-times higher than the number of intact ORF fragments identified in Australian dasyuromorphian marsupials. An additional PCR screen was done for the human genome where LINE-1 are retrotranspositionally active (Brouha et al. 2003). Specific primers were designed from the human L1HS sequence and used to amplify 100 random LINE-1 sequences, of which 40 copies were without stop codons. For all investigated species the overall mean distance was calculated giving the number of substitutions per site (p) as an average overall sequence pairs (table 1). In human and opossum, two species with known retrotranspositional activity, the results are almost identical ( $P=0.05$ ), despite belonging to two mammalian infraclasses that have been separated for 160 Ma. This is clearly different from the dasyuromorphian marsupials (table 1), where the observed mean distance for the Tasmanian devil LINE-1 alignment is  $P=0.142$ . For quoll and dunnart the overall mean distance for 100 clones is  $P=0.154$  and  $P=0.148$ , respectively (table 1), which is similar to Tasmanian devil. The number of differences between sequences from retrotranspositionally active versus inactive species is graphically shown as heatmaps (fig. 2B) where the similarity of the sequences is shown as a pairwise comparison for Tasmanian devil, dunnart, and opossum. Thereby the opossum has a much higher sequence similarity among the sequenced clones than Tasmanian devil or dunnart (fig. 2B).

The experimental screen indicates that all three dasyuromorphian species have relatively few LINE-1 sequences that can be translated into the corresponding 191 aa without stop codons and there is a high nucleotide variation among them. However, the phylogenetic tree of dasyuromorphian LINE-1 sequences (fig. 2A), shows LINE-1 evolution that is specific to the respective subfamilies, which would only be possible if



**Fig. 2.**—Results of LINE-1 sequence analyses. (A) Neighbor joining cladogram of 400 LINE-1 sequences obtained from Tasmanian devil (blue), Northern quoll (green) and dunnart (orange). The cladogram shows that the diversity of LINE-1 copies present in the respective genomes is mostly species-specific, with few common clusters from ancestral insertions. (B) Heat maps for experimental LINE-1 sequences of the dunnart, Tasmanian devil and opossum showing the sequence similarity of the experimental LINE-1 sequences for each species. (C) Boxplot of the experimentally obtained sequences queried against the respective genomes for human, opossum and the Tasmanian devil. The results for the Tasmanian devil indicate that only few of the obtained LINE-1 sequences have a good match to the genome.

LINE-1s were active for at least some time after the divergence of the subfamilies.

A PCR screen of functional LINE-1 copies in the genome of an Australian marsupial order that is closely related to *Dasyuromorphia*, the eastern barred bandicoot (*Perameles gunnii*) (order *Peramelemorphia*), yielded an ambiguous result. Previous analysis of phylogenetically informative SINE insertions could identify four WSINE1 insertions that occurred in the ancestor of *Perameles* and its sister species *Isoodon* (Gallus et al. 2015), which diverged 9 Ma (Westerman et al.

2012). After random sequencing of 100 clones, we could only recover eight LINE-1 copies that could be translated, with an overall mean group distance of 0.243 (table 1). Ongoing work on Australian kangaroo (*Macropodidae*) phylogeny have uncovered yet another possible LINE-1-silencing event, supported by a lack of phylogenetically informative LINE or SINE insertions as well as very limited amounts of intact LINE-1 copies from an experimental screen (Dodt et al. in preparation). Thus, the three investigated Australian marsupial orders (*Dasyuromorphia*, *Peramelemorphia*, *Diprotodontia*) have very similar LINE-1 inactivation patterns, with few LINE-1 fragments

that can be translated and with larger nucleotide variation than in retrotranspositionally active species. Further verification using high quality genome assemblies will be necessary to determine whether the observed LINE-1 depletion is a bias in the screening or a genuine trend among the Australian marsupial orders.

### Genome Assemblies from Short Read Data Impact the Accuracy of LINE-1 Sequences

Most of the experimentally obtained LINE-1 sequences should be detectable in the respective reference genome assemblies of Tasmanian devil, opossum or human. Using the experimental LINE-1 sequences as a query against genome assemblies (fig. 2C), we found the human and opossum LINE-1 sequences to have the greatest similarity to their respective genome assemblies. However, the result is very different when doing the same analyses for the Tasmanian devil: only *one* of the 200 experimentally sequenced LINE-1 copies is found in the assembly with 100% similarity. It is known that long multi-copy TEs, like LINE-1 sequences, are difficult to assemble (Gentles et al. 2007; Treangen and Salzberg 2012). In terms of long repetitive sequences the Sanger sequenced opossum genome (Gentles et al. 2007; Mikkelsen et al. 2007) has a better assembled genome than the Illumina sequenced Tasmanian devil genome (Murchison et al. 2012) ([supplementary table S2, Supplementary Material](#) online). The difficulty to identify Sanger sequenced LINE-1s in the Illumina sequenced Tasmanian devil genome, is likely to be typical for any genome-assembly with a large fraction of long, repetitive DNA sequences that is based on short-read sequencing technology. Thus it is crucial for any functional conclusion of autonomous TEs in a genome to take the sequencing method and assembly into account.

## Materials and Methods

### Primer Design

PCR primers were designed in the reverse transcriptase domain of LINE-1 ORF2, specifically targeting the region crucial for LINE-1 retrotranspositional activity (Cantrell et al. 2000). The forward primer is located in the homology domain 5 (including the Y/FXDD box), while the reverse primer is located outside the reverse transcriptase domain in a conserved region (Cantrell et al. 2000). The same region has been used in previous studies and covers ~573 nt (with primers 612 nt) of the LINE-1 (Cantrell et al. 2000, 2008; Grahn et al. 2005). Primer pairs were designed according to the potentially youngest LINE-1 elements deposited in RepBase (Jurka et al. 2005) from Gray short-tailed opossum (L1-1\_MD) (*M. domestica*), Tammar wallaby (L1-1\_ME) (*Macropus eugenii*), and Tasmanian devil (L1-1\_SH) (*S. harrisi*). The primers for human LINE-1 was designed according to the youngest human LINE-1 element (L1HS) deposited in Repbase

(Jurka et al. 2005). The primer positions within LINE-1 are shown in figure 1. The species-specific primers for the LINE-1 ORF2 fragment for Tasmanian devil (5' CTCTTGCGATGATGATG 3' and 5' ACCTAACCTATTCCACTGATC 3') were used for amplifying Tasmanian devil, *Dasyurus*, *Sminthopsis*, and *Perameles*. The primers (5' CTCTTGCGGATGACATGATG 3' and 5' CCCCAGTCTATTCCACTGATC 3') were used to amplify opossum and the primers (5' CTGTTTGCAGACGACATG ATT 3' and 5' CTCTGTTCTGTTCCATTGATC 3') were used for human.

### Primer Testing in Stripe-Faced Dunnart

The striped-faced dunnart (*S. macroura*) is the only representative of the genus *Sminthopsis* that has had parts of its genome sequenced. The NR and HTGS databases were screened using the Tasmanian Devil LINE-1 nucleotide sequence and the protein sequences of ORF1 and ORF2 of gray short-tailed opossum using BLASTN and TBLASTN. For the LINE-1 sequence we observed matches in 19 sequences that were retrieved from Genbank (DQ518244.1, AB326405.1, AL731834.2, AC123968.2, AL731837.2, AL731846.3, AC123538.3, AC123967.3, BX649239.1, BX649240.1, AC146781.1, BX649259.2, BX649465.2, BX649270.3, BX649310.6, BX649374.3, AC148753.1, AC148754.1, and AC186006.1). First, ORF2 sequences were extracted and manually inspected for completeness. Then, the primer sequences used for amplification of ORF2 in Tasmanian devil were blasted against the retrieved sequences (blastn, word\_size 7). To find putative primer binding sites, all hits with at least 80% identity and a minimum length of 15 nt were considered ([supplementary table S3, Supplementary Material](#) online).

### DNA Isolation, PCR Amplification, and Cloning

Total DNA was isolated following a standard phenol chloroform extraction with subsequent ethanol precipitation (Sambrook and Russell 2001). For each species ~300 ng genomic DNA was used as template in a 50 µl PCR reaction.

The ~573 nt long LINE1 ORF2 fragment was amplified using a standard PCR protocol with ExTaq (TaKaRa) to produce PCR-products with 3' adenine (A) overhangs, using stringent touch down conditions (95°denaturation/annealing<sub>1</sub> (TD, decreasing 1 degree per cycle) 65–55 °C, annealing<sub>2</sub> (24 cycles) 55 °C, 72 °C elongation, 30 cycles). Amplicons were purified with SIGMA Gen Elute PCR purification Kit (SIGMA-Aldrich) and ligated into a Topo-isomerase plasmid vector (pCR<sup>TM</sup>4-TOPO) using TA-cloning. Heat-shock transformation of *Escherichia coli* (One Shot<sup>®</sup> TOP10) chemically competent cells followed manufacturers protocol (TOPO<sup>®</sup>TA Cloning<sup>®</sup> Kit for Sequencing, Invitrogen, Life technologies, Thermo Fisher Scientific, Schwerte). Transformants were plated on LB plates containing Kanamycin (50 µg/ml). White colonies were randomly selected and PCR amplified with a



standard colony PCR protocol with VWR Taq (VWR, Darmstadt, Germany) using M13 primers. Amplicons were sequenced using Sanger sequencing using the M13 primers in both orientations with the Big Dye Termination sequencing kit (Applied Biosystems). For each species at least 100 colonies were picked and sequenced. If a sequencing reaction failed, this was removed from the data set and additional colonies were picked and sequenced from the same plated transformation reaction to generate at least 100 sequences for each species.

### Distance Estimation and Phylogenetic Analysis

The quality of all sequences was inspected manually in Geneious (Biomatters Ltd.). Poor quality sequences were removed from the analysis, and primer sequences were trimmed prior to analysis resulting in a final fragment length of about 573 nt. For each species, all sequences were aligned and mapped against the youngest LINE-1 sequence (i.e. L1-1\_MD/L1-1\_SH/L1HS). The within mean nucleotide distance for each species was calculated in MEGA7 (Kumar et al. 2016). An across species alignment was generated for the dasyuromorphian species (Tasmanian devil, *D. geoffroyi*, and *S. crassicaudata*) to gain insight into the evolution of LINE-1 in the marsupial order. Sequences were aligned with Geneious (Biomatters Ltd.) using the Geneious alignment algorithm with default parameters (65% similarity) to create a first alignment. The alignment was subsequently inspected manually and modified if necessary. Distance matrices were calculated in Geneious. Overall mean distance of each alignment was calculated with MEGA7 (Kumar et al. 2016) involving all sequenced clones for each species, and the analyses were performed with the Tamura–Nei model (Tamura and Nei 1993) and site coverage cut-off value of 95%. ML trees were calculated using MEGA7 and the General Time Reversible (GTR) evolutionary model as suggested by the model selection algorithm implemented in MEGA7, using a final alignment of three species with 400 sequences and 531 sites. A site coverage cutoff of 95% (partial deletion) was chosen. Phylogenetic trees were calculated for opossum and human as described above.

### Comparison Between Experimental LINE-1 Data and Genome Assembly

In an attempt to find the original genomic locus of the experimentally extracted LINE-1 sequence, each one of the sequences were used as a search query against the respective genome assembly (human: hg38; opossum: monDom5; Tasmanian devil: WTSI\_Devil\_ref v7.0) using BLASTN (Altschul et al. 1990). The best hit for each sequence was determined by sorting for lowest *E* value, highest length, bit-score, and identity. For Tasmanian devil the 19 clones with intact ORFs were separately investigated and screened in the published Tasmanian devil genome (WTSI\_Devil\_ref v7.0).

Each region that gave a match was extracted together with 6,000 nt up and downstream of the best match. The extracted region was screened for 1) complete LINE-1 sequence and 2) stop codons in the ORFs. Similarities among LINE-1 sequences were calculated per species in Geneious and plotted as heat-maps using R (R CoreTeam 2016).

## Supplementary Material

Supplementary tables S1–S3 and figures S1–S4 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

We thank William G. Dodt, Queensland University of Technology, and Gerrit Knapp for help with the experimental work. Axel Janke read and commented on the manuscript. This work was funded by Senckenberg, a Leibniz Institute supported by the Leibniz Association.

## Literature Cited

- Alkan C, Sajjadian S, Eichler EE. 2011. Limitations of next-generation genome sequence assembly. *Nat Methods*. 8:61–65.
- Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.
- Boissinot S, Roos C, Furano AV. 2004. Different rates of LINE-1 (L1) retrotransposon amplification and evolution in new world monkeys. *J Mol Evol*. 58:122–130.
- Brouha B, et al. 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A*. 100:280–285.
- Cantrell MA, Grahn RA, Scott L, Wichman HA. 2000. Isolation of markers from recently transposed LINE-1 retrotransposons. *BioTechniques* 29:1310–1316.
- Cantrell MA, Scott L, Brown CJ, Martinez AR, Wichman HA. 2008. Loss of LINE-1 activity in the megabats. *Genetics* 178:393–404.
- Casavant NC, et al. 2000. The end of the LINE?: lack of recent L1 activity in a group of South American rodents. *Genetics* 154:1809–1817.
- Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol*. 7:567–580.
- Feng Q, Moran JV, Kazazian HH, Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87:905–916.
- Gallus S, et al. 2015. Evolutionary histories of transposable elements in the genome of the largest living marsupial carnivore, the Tasmanian devil. *Mol Biol Evol*. 32:1268–1283.
- Gallus S, Janke A, Kumar V, Nilsson MA. 2015. Disentangling the relationship of the Australian marsupial orders using retrotransposon and evolutionary network analyses. *Genome Biol Evol*. 7:985–992.
- Gentles AJ, et al. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res*. 17:992–1004.
- González J, Petrov DA. 2009. The adaptive role of transposable elements in the drosophila genome. *Gene* 448:124–133.
- Grahn RA, Rinehart TA, Cantrell MA, Wichman HA. 2005. Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. *Cytogenet Genome Res*. 110:407–415.
- Gu W, et al. 2007. SINEs, evolution and genome structure in the opossum. *Gene* 396:46–58.

- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, and Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110:462–467.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33:1870–1874. doi: 10.1093/molbev/msw054
- Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol.* 16:793–805.
- Meredith RW, et al. 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521–524.
- Mikkelsen TS, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167–177.
- Mills RE, Bennett EA, Iskow RC, Devine SE. 2007. Which transposable elements are active in the human genome? *Trends Genet.* 23:183–191.
- Murchison EP, et al. 2012. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* 148:780–791.
- Nilsson MA. 2016. The devil is in the details: transposable element analysis of the Tasmanian devil genome. *Mob Genet Elements.* 6:e1119926.
- Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3' sequences: a review. *Gene* 205:229–243.
- Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *BioEssays* 31:703–714.
- Platt RN, Ray DA. 2012. A non-LTR retroelement extinction in *Spermophilus tridecemlineatus*. *Gene* 500:47–53.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [cited Apr 4, 2016]. Available at: <https://www.R-project.org/>.
- Rinehart TA, Grahn RA, Wichman HA. 2005. SINE extinction preceded LINE extinction in Sigmodontine rodents: implications for Retrotranspositional dynamics and mechanisms. *Cytogenet Genome Res.* 110:416–425.
- Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Scott L, Kuroiwa A, Matsuda Y, Wichman HA. 2006. X accumulation of LINE-1 Retrotransposons in *Tokudaia osimensis*, a spiny rat with the Karyotype XO. *Cytogenet Genome Res.* 112:261–269.
- Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with Retroposons. *Trends Ecol Evol.* 19:545–553.
- Stewart C, et al. 2011. A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* 7:e1002236.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10:512–526.
- Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.* 13:36–46.
- Westerman M, et al. 2012. Phylogenetic relationships of living and recently extinct bandicoots based on nuclear and mitochondrial DNA sequences. *Mol Phylogenet Evol.* 62:97–108.

Associate editor: Ellen Pritham