Molecular Phylogenetics and Evolution 168 (2022) 107393

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

A genomic survey of LINE elements in Pipidae aquatic frogs shed light on Rex-elements evolution in these genomes

Joana Moura Gama [a,b], Adriana Ludwig [c,*], Camilla Borges Gazolla [a,b], Dieval Guizelini [d], Shirlei Maria Recco-Pimentel [e], Daniel Pacheco Bruschi [a,b,*]

[a] Programa de Pós-Graduação em Genética (PPG-GEN), Universidade Federal do Paraná (UFPR), Curitiba, Brazil
[b] Laboratório de Citogenética evolutiva e Conservação Animal (LabCeca), Departamento de Genética, Universidade Federal do Paraná (UFPR), Brazil
[c] Laboratório de Ciências e Tecnologias Aplicadas em Saúde (LaCTAS), Instituto Carlos Chagas, Fiocruz-PR, Brazil
[d] Programa de Pós-Graduação em Bioinformática, Universidade Federal do Paraná, Curitiba, Brazil
[e] Departamento de Biologia Estrutural e Funcional, Universidade Estadual de Campinas (UNICAMP), Brazil

ARTICLE INFO

Keywords:
Transposable elements
LINEs
Horizontal transfer
Anura

ABSTRACT

The transposable elements (TE) represent a large portion of anuran genomes that act as components of genetic diversification. The LINE order of retrotransposons is among the most representative and diverse TEs and is poorly investigated in anurans. Here we explored the LINE diversity with an emphasis on the elements generically called Rex in Pipidae species, more specifically, in the genomes of *Xenopus tropicalis*, used as a model genome in the study of anurans, the allotetraploid sister species *Xenopus laevis* and the American species *Pipa carvalhoi*. We were able to identify a great diversity of LINEs from five clades, Rex1, L2, CR1, L1 and Tx1, in these three species, and the RTE clade was lost in *X. tropicalis*. It is clear that elements classified as Rex are distributed in distinct clades. The evolutionary pattern of Rex1 elements denote a complex evolution with independent losses of families and some horizontal transfer events between fishes and amphibians which were supported not only by the phylogenetic inconsistencies but also by the very low Ks values found for the TE sequences. The data obtained here update the knowledge of the LINEs diversity in *X. laevis* and represent the first study of TEs in *P. carvalhoi*.

1. Introduction

Transposable elements (TEs) are ubiquitous genomic components and correspond to a significant portion of the eukaryotic chromosomes, acting as an important evolutionary substrate to genomic diversification (Biémont and Vieira, 2006; Farré et al., 2011; Wells and Feschotte, 2020). Although still restricted to a few species, the analysis of the repetitive fraction in the anuran genomes is no exception to the rule of other vertebrates (Chalopin et al., 2015; Feschotte and Pritham, 2007; Kaminker et al., 2002; Keinath et al., 2015; Sotero-Caio et al., 2017), with massive invasion and expressive diversity of TEs (Chalopin et al., 2015; Hellsten et al., 2010; Sun et al., 2015).

As natural components of genomes, the TEs are vertically transferred to the derived lineages. Additionally, numerous of evidence has revealed the role of the horizontal transfer (HT) as a second mechanism responsible for the spreading of these sequences among the genomes, being able to transfer these elements between reproductively isolated species (Silva and Kidwell, 2000; Wallau et al., 2018). Most of the HT described for vertebrates involve ray-finned fish (93.7%) and only 3% involve mammals and birds (Zhang et al., 2020).

Retrotransposons (Class I elements) compose a large portion of the TEs in most genomes (Chalopin et al., 2015). Due to their transposition, by reverse transcription mechanism, they are identified as an important genomic component that contributes (directly or indirectly) to the increase in the size of some genomes (Elliott and Gregory, 2015; Kidwell, 2002). Among retrotransposons, the LINE (*Long Interspersed Nuclear Element*) order (Wicker et al., 2007) stands out mainly for being present in practically all eukaryotic organisms with a great diversity of clades and families (Chalopin et al., 2015).

LINE elements are characterized by the presence of at least one open reading frame (ORF), the *Pol*-like, encoding reverse transcriptase (RT) and endonuclease (EN) domains (Eickbush and Malik, 2002). Some LINEs also have an additional upstream ORF called *Gag*-like (Jurka et al., 2007; López-Flores and Garrido-Ramos, 2012). Moreover, it is possible to find a 3′-end with tandem repeats or regions rich in adenosine (A) (Volff et al., 2001, 2000, 1999; Wicker et al., 2007).

* Corresponding authors at: Programa de Pós-Graduação em Genética (PPG-GEN), Universidade Federal do Paraná (UFPR), Curitiba, Brazil (D.P. Bruschi).
E-mail addresses: adriludwig@gmail.com (A. Ludwig), danielbruschi@ufpr.br (D.P. Bruschi).

https://doi.org/10.1016/j.ympev.2022.107393
Received 12 August 2021; Received in revised form 9 November 2021; Accepted 25 December 2021
Available online 17 January 2022

The classification system proposed by Wicker et al. (2007) considers the LINE order elements divided into five superfamilies: R2, L1, RTE, I and Jockey, based on Eickbush and Malik (2002). However, additional groups/clades can now be recognized and the current classification available in Repbase established 31 clades corresponding to LINE elements that together with *Penelope* and SINEs compose the so-called non-LTR retrotransposons (Kojima, 2019). These clades are organized into eight major groups: CRE group (for CRE clade), R2 group (composed by NeSL, R4, R2 and Hero clades), Dualen group (for RandI/Dualen clade), L1 group (composed by L1, Proto1 and Tx1 clades), RTE group (composed by Proto2, RTE, RTEX and RTETP clades), I group (compose by I, Nimb, Ingi, Vingi, Tad1, Loa, R1, Outcast and Jockey clades), CR1 group (composed by CR1, L2, L2A, L2B, Kiri, Rex1, Crack and Daphne clades) and Ambal group (for Ambal clade). Still, the Odin clade, established by Volff et al (2004), has not been included in any major groups since its position in the tree is unclear (Kojima, 2019).

A clear example of the taxonomic identification dilemma of TEs is the LINE-like elements generically classified/named as Rex. These elements were first identified in the bony fish *Xiphophurus maculatus*, being named according to the order in which they were discovered, such as Rex1 (Retroelement of *Xiphophurus* 1), Rex3 (Retroelement of *Xiphophurus* 3), Rex5 (Retroelement of *Xiphophurus* 5) and Rex6 (Retroelement of *Xiphophurus* 6) (Volff et al., 2001, 2000, 1999). The Rex elements are widely distributed in teleost fish (Volff et al., 2000, 1999) and initially considered specific elements of this species group (Volff et al., 2000, 1999), however, homologous sequences have also been found in other organisms, like turtles (Rex6) and alligator (Rex1) from Amazon (Noronha et al., 2016; Oliveira et al., 2021) and anurans of the *Physalaemus* (Rex1) genus (Nascimento et al., 2015). These elements are eventually treated as belonging to the same group, often referred to as the Rex family (Carducci et al., 2018; Favarato et al., 2017). However, there is no evidence of a natural grouping of these sequences. Phylogenetic reconstructions support that Rex elements are not monophyletic, being Rex1 recovered in the Rex1 clade (Volff et al., 2000), Rex3 recovered in the RTE clade (Lovšin et al., 2001), Rex5 recovered in the L2 clade (Lovšin et al., 2001) and Rex6 recovered in the R4 clade (Burke et al., 2002).

The TE repertoire of Anura genomes is underexplored, though, it is possible to observe a great diversity of LINE-like elements in *Xenopus tropicalis*, *Rhinella marina* and *Nanorana parkeri* (Edwards et al., 2018; Hellsten et al., 2010; Sun et al., 2015). Unfortunately, most of the knowledge about TEs in anuran genomes is limited to the RepeatMasker annotation during the genome assembly process. *Xenopus tropicalis* is a representative species of the family Pipidae and has been used as a reference genome for Anura in several genomic comparative projects of vertebrates.

The Pipidae family is an interesting evolutionary model due to their complex morphological and ecological traits, being highly specialized to an aquatic lifestyle, and by their biogeographical diversification context (Cannatella, 2015; Irisarri et al., 2011). The Pipidae family includes 41 species from four genera, three distributed in Sub-Saharam Africa (genera *Hymenochirus*, *Pseudhymenochirus* and *Xenopus*) and one restricted to Central and South America (genus *Pipa*). The genus *Pipa* is a sister group of the African lineages [*Xenopus* + (*Hymenochirus*, *Pseudhymenochirus*)] and their diversification is still the subject of controversy (Cannatella, 2015). Estimates suggest, for example, at least 100 million years (My) the age of the last common ancestor of *X. tropicalis* and *Pipa carvalhoi* (Cannatella, 2015) and 48 My the last common ancestor of *X. tropicalis* and *X. laevis* (Cannatella, 2015; Session et al., 2016). It is interesting to mention that *X. laevis* has an allotetraploid origin, formed by subgenomes (nominated L and S) that diverged from each other around 34 My ago (Mya) (Session et al., 2016).

Here, we designed genomic searches on *X. tropicalis*, *X. laevis,* and *P. carvalhoi* genomes to access the content of sequences homologous to the LINE elements generically called Rex (Rex1, Rex 3, Rex 5 and Rex 6), aiming to contribute on (i) their diversity, (ii) classification and (iii) evolutionary dynamics in each genome. Our data reveals greater diversity of LINEs in Pipidae, also confirm that elements denominated Rex are distributed in different clades and has a complex evolution with several HT events among fishes and Pipidae.

2. Material and methods

2.1. Pipa carvalhoi genome

We used the draft version of *P. carvalhoi* genome, recently sequenced by the Laboratório de Citogenética Evolutiva e Conservação Animal at Universidade Federal do Paraná (UFPR), Brazil. The genome was sequenced on the Illumina HiSeq-400 platform, 46x coverage. The raw reads were trimmed with Trimmomatic (Bolger et al., 2014) and assembled with Velvet 1.2.10 (Zerbino and Birney, 2008) for de novo assembly, with value 105 to parameter for hash length (kmer) and it is currently at a level of contigs (2.4 Gpb, 5,029,551 contigs). The hardware used in the assembly is SGI UV100 systems, it has 64-cores in 8 processors Intel (R) Xeon (R) CPU E5-4640 0 @ 2.40 GHz and with 512 Gb of memory. The genome assembly is available at http://200.236.3.93/xeno/velvet/contigs_k105.fa. Quast analysis reports (Gurevich et al., 2013) indicate that the assembly achieved 830,139 contigs greater than 500 bp and the values for N50, N75 are 1,968 and 914, and for metrics, L50 and L75 are 145,457 and 375,377 and the GC context is 38.16 (Supplementary File 1). BUSCO (Seppey et al., 2019) was used to evaluate the completeness of the genome. The high fragmentation of genome assembly is reflected in the high proportion of genes recovered as “fragmented” and missing” (Supplementary File 2). For this work, the contigs were filtered by size keeping only those larger than 400 bp (1,153,690 contigs, 1.3 Gpb of sequences).

2.2. Analysis of Rex elements and the LINE diversity in Pipidae

To understand the diversity of retrotransposons generically called Rex in Pipidae, we performed searches for homologous sequences in *X. tropicalis, X. laevis* and *P. carvalhoi* genomes using tBLASTn (Altschul et al., 1997). Until July 2019, nine Rex sequences were available in the Repbase for *X. tropicalis* (named REX1-1_XT to REX1-9_XT), while none was described for *X. laevis*. The Rex families from *X. tropicalis* families with preserved coding capacity (REX1-1_XT to REX1-5_XT and REX1-9_XT) and those from fish species *Xi. maculatus* (REX1 Accession number AF155728.1, REX3 and REX5) and *Takifugu rubripes* (REX6) had their RT domain isolated and used as queries in local tBLASTn against *P. carvalhoi* genome. To extend searches of closely related sequences in *X. tropicalis* and *X. laevis*, the sequences previously retrieved in *P. carvalhoi* were filtered by identity (0.65) using CD-hit (Huang et al., 2010) and their complete RT domains were used as queries in online tBLASTn against *X. tropicalis* (GCA_000004195.4) and *X. laevis* (GCA_001663975.1) genomes. For each species, the 10 hits with greater significance were analyzed (considering the highest score and *e-value* $< 10^{-4}$), recovering the sequences together with around 1 kb of flanking regions. Redundancy was eliminated manually. For each hit, the detection of ORFs was performed by the ORFinder tool (https://www.ncbi.nlm.nih.gov/orffinder/) and protein domains were predicted using the NCBI CD-Search (Marchler-bauer and Bryant, 2004), both with default parameters.

Complete or nearly complete RT domains (greater than 70% of domain size) were isolated to conduct phylogenetic analyzes. To better visualize the LINE diversity in *Xenopus* and the relationship among Rex elements, all *Xenopus* LINE sequences (108 families, 98 from *X. tropicalis* and 10 from *X. laevis*) deposited in the Repbase database version 23.11 (Jurka et al., 2005) were also added to the matrix (Supplementary File 3). LINE sequences known to belong to distinct clades (Eickbush and Malik, 2002; Kojima, 2019; Malik et al., 1999; Volff et al., 2004) were included in the matrix as diagnostic sequences of each major clade (Supplementary File 4). In-house Python scripts were used to isolate the

domains and manage the sequences.

The RT amino acid sequences were aligned using the MAFFT platform version 7 (https://mafft.cbrc.jp/alignment/software/) using the option "L-INS-i" (Katoh et al., 2017), visualized in GeneDoc 2.7 (Nicholas and Nicholas, 1997) and inspected manually. The trimAl tool (Capella-gutiérrez et al., 2009), implemented in the NGPhylogeny.fr service (Lemoine et al., 2019), was used to trim the alignment with the following parameters: GAP threshold 0.9, similarly threshold 0.0 and consistency threshold 60. The alignment is available in Supplementary File 5. Two methods were used to infer the evolutionary trees: (i) The maximum likelihood (ML) by the RAxML-HPC BlackBox tool (Kozlov et al., 2019) implemented on the CIPRES platform (Miller et al., 2010) using the LG + G + I as suggested by the Bayesian Information Criterion (BIC) of the model test implemented in MEGA X (Kumar et al., 2018); and (ii) Bayesian analysis (BA) was inferred under a mixed model, performing 5,000,000 generations of Markov Monte Carlo chain (MCMC) and sampling trees every 1,000 generations. The 25% of the initial results (burn-in) were discarded and the rest of the trees were summarized in a consensus.

The trees were visualized and edited in ITOL (Interactive Tree of Life - https://itol.embl.de/) (Letunic and Bork, 2019) and rooted by the canonical sequence of SLAC element from the CRE clade, as it is the first branched lineage of LINEs (Eickbush and Malik, 2002; Malik et al., 1999)

2.3. Characterization of Rex1 clade in Pipidae

We used the RT nucleotide portion of REX1-1 to REX1-9 to search for homologous sequences in *X. tropicalis*, *X. laevis* and *P. carvalhoi* genomes by tBLASTx. We also proceeded the same search in four fish genomes, *Danio rerio* (GCA_000002035.4), *Erpetoichthys calabaricus* (GCA_900747795.2), *Simochromis diagramma* (GCA_900408965.1) and *Denticeps clupeoides* (GCA_900700375.2). Hits with an alignment size corresponding to 70% long and at least 50% identity in amino acid level were retrieved from genomes and the redundancy was eliminated using cd-hit-est (identity 1.0). For the evolutionary tree inference, all sequences were aligned using MAFFT and the alignment was trimmed as described above. The alignment is available in Supplementary File 6. A Bayesian tree (parameters lset nst = 6 rates = gamma) was inferred as described above performing 60,000,000 MCMC generations.

2.4. Evaluating horizontal transfer hypothesis

Possible events of HT were investigated between an RTE element in *Helobdella robusta* and *P. carvalhoi* and among Rex1 families from Pipidae and fish species. We tested the hypothesis by a widely used approach (Bartolomé et al., 2009; Ludwig et al., 2008; Silva and Kidwell, 2000), comparing the synonym substitution rate (Ks) (a measure of neutral evolution) found for the retroelement and for orthologous genes from the species involved. Lower Ks for the TEs than genes reflects the lower divergence time of sequences compared to the species divergence time, suggesting HT (Ludwig et al., 2008; Silva and Kidwell, 2000).

A total of 40 single-copy ortholog genes were chosen for the analysis. We used the proteome of *X. tropicalis* to search homologous sequences in *P. carvalhoi*, by tBLASTn against the *P. carvalhoi* coding sequences (excluding sequences shorter than 600 bp) provided by BUSCO. All sequences identified were then evaluated for the number of copies and the presence of orthologs in *X. laevis, H. robusta,* and the four fish genomes totalizing 32 genes. Another 8 genes were chosen using OrthoDB tool (https://www.orthodb.org/) and the coding sequence of these genes in *P. carvalhoi* was assembled manually, using the *X. tropicalis* protein counterpart as a query in tBLASTn to retrieve the contigs containing the exons. The information of the genes is available in Supplementary File 7. For the retrotransposons, one copy of each species possibly involved in HT was used for this analysis.

The MUSCLE tool was used to obtain the codon alignment of genes and TEs in the MEGA X program after the translation of nucleotide sequences. Some TE sequences were slightly edited to recover the coding region. The Ks value of each gene and TEs between species was calculated using an improved LPB93 method (Li, 1993; Pamilo and Bianchi, 1993) estimated in DAMBE tool (Data Analysis in Molecular Biology and Evolution) (Xia, 2018). Taking into account that the preferential use of codons has a direct effect on Ks, we estimate the use of codons by calculating the NC (effective number of codons) also in the DAMBE tool. The Kimura 2 parameter (k2p) distance (Kimura, 1980) between all copies was estimated with the distmat application from EMBOSS (Rice et al., 2000). The histogram distributions of distances were plotted with Python Matplotlib-v3.3.2 (Hunter, 2007) and edited in Inkscape software. Possible HT events between the fish species were not explored here.

3. Results

3.1. Isolation of sequences homologous to Rex elements and other LINEs in P. carvalhoi

Searches for homologous sequences to the elements nominated as Rex in the *P. carvalhoi* genome retrieved copies with a variety in the level of integrity. In total, 181 contigs were analyzed and the majority presented the RT domain (44 of which were complete domain and 91 were incomplete or truncated). See Supplementary File 8 for details of copies and Supplementary File 9 to access the sequences. As expected for the approach of search that we used, we recovered also more divergent sequences that belong to LINE families/groups other than Rex as we confirmed by the evolutionary tree (see next section). Only 18 of these contigs had both RT and EN domains (seven in the same ORF and 11 in separate ORFs). Three contigs presented only the EN domain. One characteristic of LINE elements is the presence of these two domains in a single ORF and are often truncated at their 5′end due to the integration of prematurely terminated reverse transcripts (Jurka et al., 2007; Malik et al., 1999; Volff et al., 2000). Thus, the absence of the EN domain in most copies could be the result of truncation in the integration process, however, several contigs that we analyzed are short and it was not possible to evaluate the 5′ region. Additionally, the presence of both domains in separate ORFs indicates that these copies identified in *P. carvalhoi* are possibly degenerate.

For several confirmed Rex1 copies a domain of unknown function (DUF1891 domain) is predicted after the RT domain (Supplementary File 10). It was not possible to observe structures of microsatellites in the 3′-terminal region of the elements as it was observed for Rex1 and Rex3 of fish (Volff et al. 1999, 2000) and in the different Rex1 families of *X. tropicalis* (as information provided in the Repbase). These repetitive regions may not be present in the *P. carvalhoi* elements or they were not assembled correctly.

For some copies (pipa1240930, pipa1505315, pipa117387 and pipa445583) it is possible to observe the presence of an additional ORF1 upstream the EN/RT ORF. In the ORF1 product, it was predicted an AIR1 domain (Arginine methyltransferase interaction protein, RING Zn-finger) in the same regions as it was predicted a PTZ00368 superfamily domain (universal mini-circle sequence binding protein -UMSBP). AIR1 is also found in Het-A and TAHRE, two *Drosophila* elements from Jockey group (Saint-Leandre et al., 2019). Moreover, in almost all these copies, a called zf-RVT domain (zinc-binding in reverse transcriptase) was predicted downstream of the RT. This C-terminal zf domain is found in Tx1 and other LINEs (Lu et al., 2020). The copy pipa1240930 is putative complete and presents a poly-A tail end. Poly-A tail is also found in the pipa117387, but the second ORF is truncated. The evolutionary tree (see next section) indicates these copies belong to the Tx1 clade.

3.2. LINE families diversity in Pipidae

To understand the classification of the LINE copies recovered from

P. carvalhoi genome and their relationship with other LINEs, we reconstructed an evolutionary tree of the RT domain. Besides the LINE copies from *P. carvalhoi*, all LINE elements from *X. tropicalis* and *X. laevis* described in the Repbase were included together with LINE sequences from other organisms of major known distinct LINE clades. Moreover, we included LINE copies from *X. tropicalis* and *X. laevis* after a genomic search using *P. carvalhoi* divergent copies (contigs pipa402737, pipa72182, pipa205717, pipa456686, pipa84507, pipa445583) as queries. In these genomic searches, we recovered a total of 60 and 56 sequences in *X. tropicalis* and *X. laevis* genomes, respectively, and those sequences with complete or nearly complete RT domain were included in the tree.

The evolutionary trees based on BA (Fig. 1) and ML (Supplementary File 11) allowed us to understand the classification and both phylogenetic methods recovered similar sequences-clusters, while the BA topology exhibits better resolution among the LINEs groups previously designed by Kojima (2019). Thus, the evolutionary relationships among sequences described here are based on the BA consensus tree. We recognized two well-supported monophyletic arrangement that grouped (i) sequences from clades/superfamilies (Tad1, LOA, R1) + (Ingi (I (Odin, Jockey, (Rex1 (CR1, L2))))) and (ii) sequences from clades/superfamilies (R2, R4, NeSL) + (RTE (L1, Tx1)).

The relationship among the TE families is well supported in general, making it possible to recover the common origin with high support for the main clades of LINEs. From the *Xenopus* canonical families available in the Repbase, we observed that five clades were found for *X. tropicalis* (Rex1, CR1, L1, L2 and Tx1) and only three clades for *X. laevis* (L1, L2 and CR1), although the genomic searches in *X. laevis* also recovered copies grouped with Rex1 and Tx1, indicating that some LINE families are missing in the *X. laevis* TE library. Moreover, some *X. laevis* copies

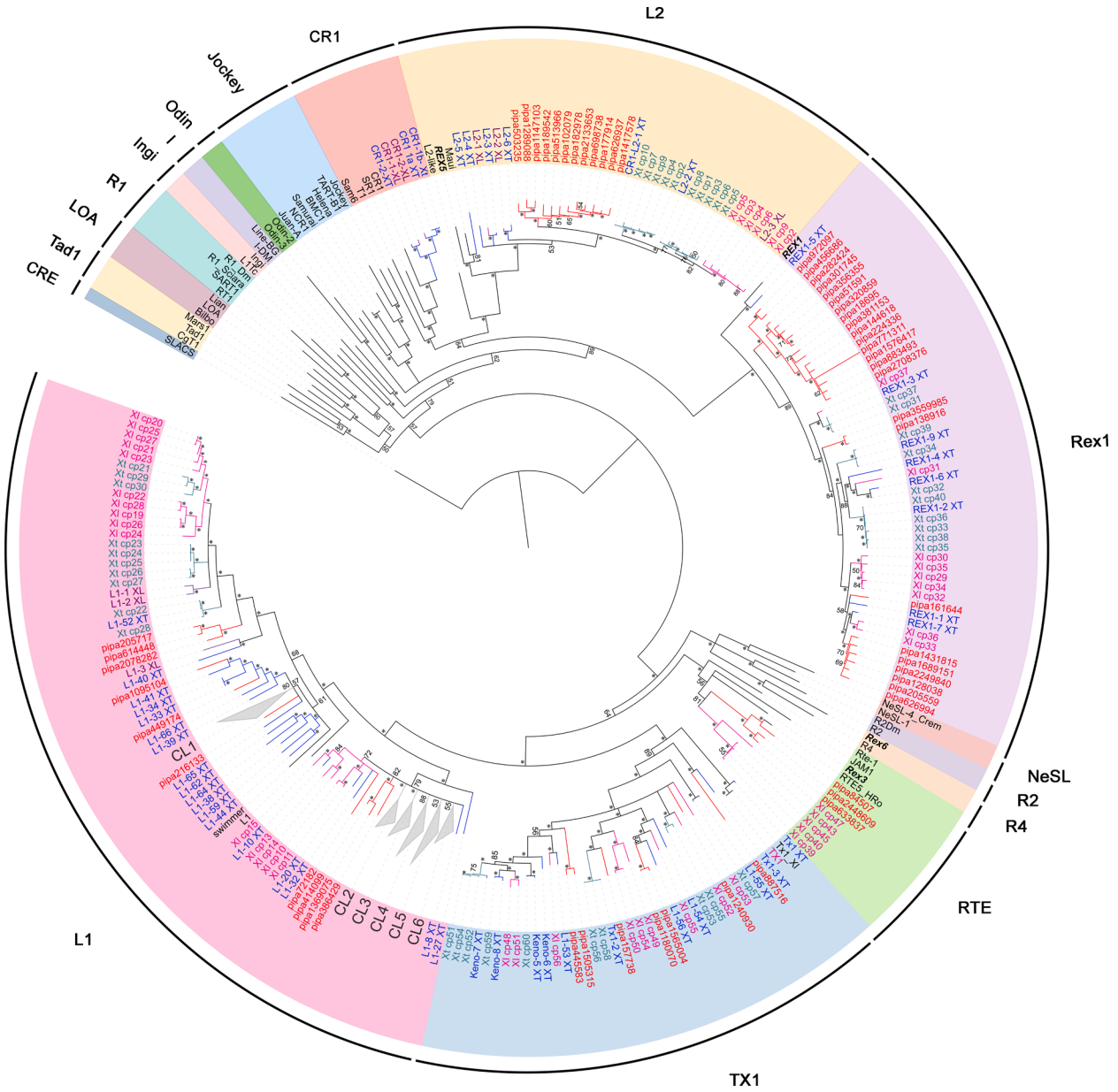

Fig. 1. BA 50% majority rule consensus phylogenetic tree of LINEs based on the RT amino acid sequence among Pipidae, rooted by CRE clade. Posterior probability higher than 90 was replaced by an asterisk (*). Copies from *P. carvalhoi* genome are in red, *X. laevis* are pink and *X. tropicalis* are turquoise-green, while copies from Repbase are dark-blue from *X. tropicalis* and purple from *X. laevis.* Some clades were collapsed into clusters of sequences represented as CL1, CL2, CL3, CL4, CL5 and CL6 whose composition is available in Supplementary File 13. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

grouped within RTE clade that is absent in *X. tropicalis* as confirmed by the specific searches. In *P. carvalhoi* genome, we recovered sequences from L1, Tx1, RTE, Rex1 and L2 clades.

Most families of the *X. tropicalis* deposited on Repbase were recovered in the respective clade accordingly to their name, with a few exceptions. Some families named as L1 are clear TX1 elements (L1-53 XT, L1-54 XT, L1-55 XT, L1-56 XT). The families named Keno-5 XT, Keno-6 XT, Keno-7 XT e Keno-8 XT, a TE family specialized in inserting in the same site of U2 small nuclear RNA gene (Kojima and Fujiwara, 2004), also belong to the Tx1 clade. Moreover, there is a family ambiguously named CR1-L2-1_XT that belongs to the L2 clade. It is evident that the elements generically called Rex and used as queries were not recovered as a single cluster but were distributed in different clades (Rex1- clade Rex1, Rex3 - clade RTE, Rex5 - clade L2 and Rex 6 - cluster R4).

The sequences retrieved in Pipidae genomes using Rex1 sequences from *X. tropicalis* as queries were all grouped in the Rex1 clade, as expected for closely related queries. On the other hand, using as query the Rex3, Rex5 and Rex6 of fish, the obtained sequences from the Pipidae species were not always recovered within the same clade as the respective query. Sequences retrieved from Rex3 search grouped in the RTE, but also in the L1 clade. All sequences retrieved from Rex5 search grouped as expected in the L2 clade while no sequences retrieved from Rex6 search were recovered with the query in the R4 clade, but they were found in the L1 and TX1 clades.

The well-supported Rex1 clade includes as a basal cluster the Rex1 sequence from *Xi. maculatus* (named REX1 in the tree) and the family REX1-5_XT from *X. tropicalis* (Fig. 1). Additionally, we recognized two major derived clusters containing Rex1 sequences previously deposited in the Repbase and copies recovered from the three Pipidae genomes which could suggest the presence of these families on the genome of the most recent common ancestor of Pipidae. However, the evolutionary history of the Rex1 elements is much more complex as detailed below (see topic 3.3).

Concerning the Rex3 element, similar sequences were detected in *X. laevis* and *P. carvalhoi* genomes, assembled in the RTE clade together with JAM1 from *Aedes aegypti*, Rte-1 from *Caenorhabditis elegans* and RTE-5_Hro from *H. robusta* (Fig. 1). No RTE sequence was recovered from *X. tropicalis* genome, as described previously (Hellsten et al., 2010) and this is the first report of RTE clade in Pipidae species. It can be suggested that this family went into a process of degeneration after the separation of *Xenopus* species and was completely lost in *X. tropicalis*. As expected, sequences from *P. carvalhoi* are closely related to *X. laevis* copies. However, unexpectedly, the element RTE-5_Hro from the leech *Helobdella robusta* was recovered inside a clade containing *P. carvalhoi* copies close to the copy pipa84507 (posterior probability: 1.0). *A priori*, this phylogenetic incongruence could be explained by (i) horizontal transfer, (ii) sequence contamination or (iii) ancestral polymorphism followed by a differential assortment of copies and/or stochastic loss aggravated by the reduced species sampling. A contamination hypothesis is less probable since several copies are found in *H. robusta* genome and this RTE is confirmed in two Pipidae species. Trying to solve this question, we tested the HT hypothesis based on the Ks value presented by the TE and the host genes. We found that the Ks value for the TE is comparable to the values found for the host genes, suggesting that the TE was inherited vertically from the common ancestor, along with the other genes. Thus, despite the phylogenetic inconsistency, there is not enough evidence to infer HT (see detailed description in Supplementary File 12).

The clade L2 encompasses the diagnostic elements L2-like from *Ciona intestinalis*, Maui from *T.rubripes*, the query Rex5 from *Xi. maculatus* and L2 elements from *X. tropicalis* and *X. laevis* previously deposited on Repbase (Fig. 1). The copies recovered from *X. laevis* and *X. tropicalis* belong to L2-3_XL and to L2-2_XT and CR1-L2-1_XT families, respectively. The *P. carvalhoi* copies grouped in one clade and may constitute one or more families of L2.

As we already mentioned, the sequences isolated from Pipidae using Rex6 from *T.rubripes* as query were not assembled with the query but were recovered on TX1 and L1 clades. This happened because we used amino acid sequences as query and consequently more divergent sequences can be recovered. As neither recovered copy from Pipidae nor their families from Repbase were recovered in the clade R4, it is possible to conclude this LINE clade is absent in Pipidae. Although TX1 and L1 clades were not the focus of our work, we can observe a great diversity of sequences for both clades shared among the three Pipidae species revealing an ancient presence of these elements on the ancestor genome of these organisms. The L1 clade is the most diverse in the number of families (67 for *X. tropicalis* and 3 for *X. laevis*) and some internal clades were collapsed in the tree (to complete list of families, see Supplementary File 13).

3.3. Evolutionary dynamics of Rex1 clade in Pipidae

The Repbase record has registered the occurrence of nine families of Rex1 in *X. tropicalis* (named REX1-1_XT to REX1-9_XT). No *X. laevis* no Rex1 family was available in the Repbase, until July 2019. The analysis of these sequences shows a considerable divergence in nucleotide (Supplementary file 14) and protein levels. They present an overall mean nucleotide divergence of 43% over around 2400 bp of core conserved alignment, being the REX1-3_XT and REX1-8_XT the closest families (24.9% of divergence). REX1-1_XT and REX1-8_XT present a deletion in the 5′ region. The structure of a large ORF containing the EN (Exo_endo_phos) and RT (RT_nLTR_like) domains is preserved in REX1-2_XT, REX1-3_XT, REX1-4_XT and REX1-6_XT. REX1-5_XT and REX1-9_XT present a structure of a large ORF but only the RT domain is predicted. As we found for *P. carvalhoi* Rex1 copies, the *X. tropicalis* Rex1 families also contain the domain of unknown function (DUF1891) predicted right after the RT domain. These families also have distinct sequence repeats in the 3′ end (see more in Supplementary File 15).

Trying to understand the dynamic of Rex1 evolution in Pipidae, we proceeded with an additional wide search for homologous sequences in *X. tropicalis*, *X. laevis* and *P. carvalhoi*. For the clustering analysis, we choose to employ nucleotide sequences to be able to use all copies retrieved and encompass those with non-preserved coding capacity. This would give a more realistic tree since several copies cannot be included in the amino acid tree. We expand the searches to some ray-finned fish genomes (Actinopterygii), *Danio rerio*, *Erpetoichthys calabaricus*, *Simochromis diagramma* and *Denticepes clupeoides* because previous online BLASTn searches (data not shown) of different *X. tropicalis* Rex1 families showed an unexpectedly high level of identity (greater than 80%) with these distantly related genomes raising the hypothesis of HT. Tetrapods evolved from lobe-finned fishes (sarcopterygians) that separate from the Actinopterygii group around 450 mya (Pyron, 2010) (Fig. 2).

Our evolutionary tree of Rex1 including sequences from Pipidae and fish shows a complex pattern of sequence grouping (Fig. 3 and Supplementary File 16). We observed species-specific amphibian groups that are positioned closer to fishes than other amphibians, such as XL-group 1, PC-group 3 and Rex1-4_XT, all related to *E. calabaricus*; and Rex1-5_XT related to *D. rerio*. In all these cases the overall high sequence

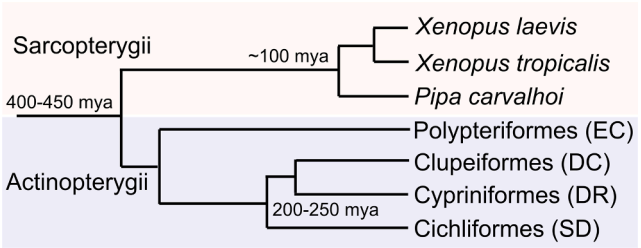

Fig. 2. Phylogenetic relationship of species used in our work. The cladogram and estimative of branching were was drawn based on previous work (Berthelot et al., 2014; Betancur-R et al., 2017; Cannatella, 2015; Hedges et al., 2015; Pyron, 2010; Safian et al., 2021; Session et al., 2016).

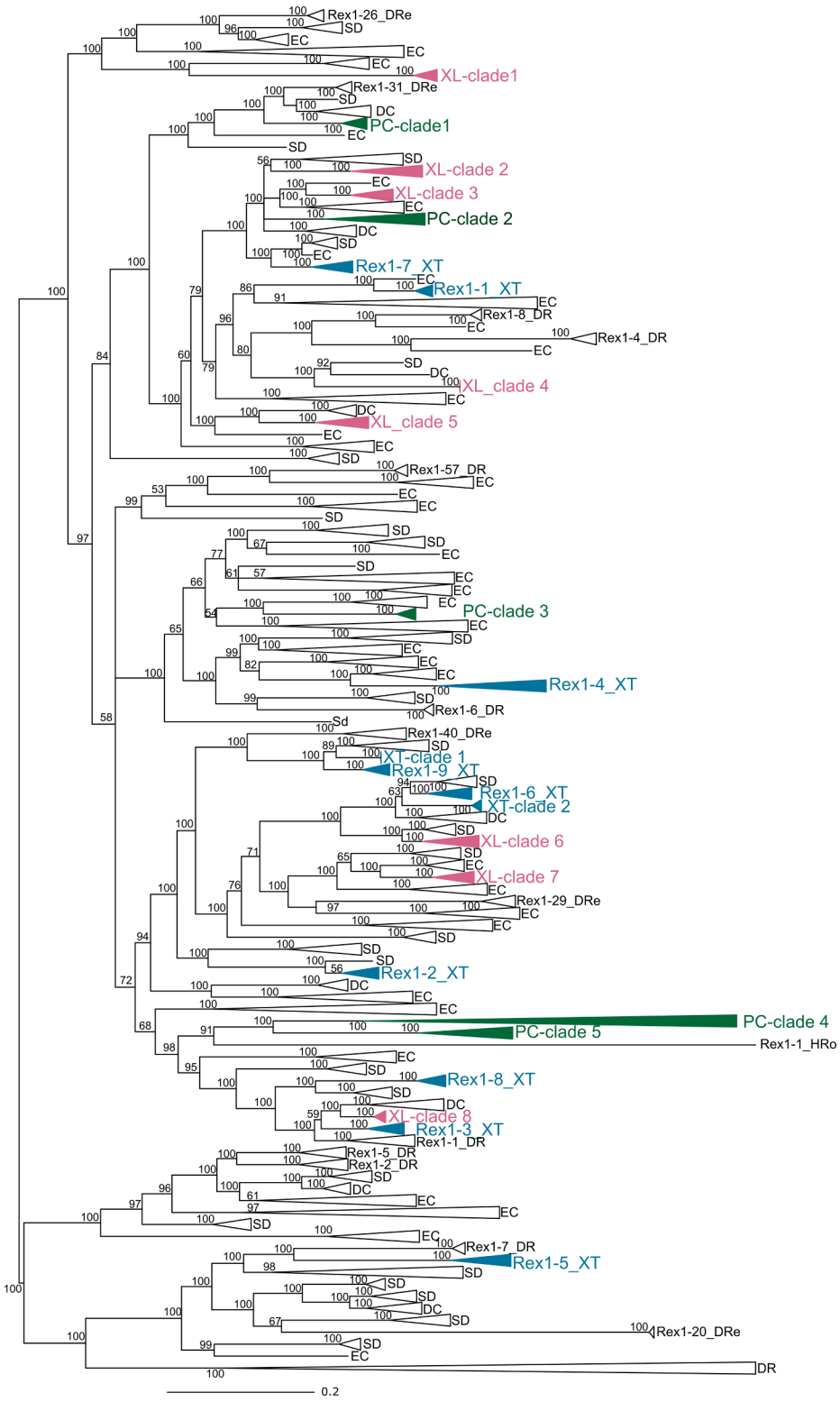

Fig. 3. BA 50% majority rule consensus phylogenetic tree of Rex1 based on RT nucleotide sequences from the three Pipidae species and four fish species. The tree was rooted by the midpoint. Posterior probability is shown near the nodes. Clades from the same species were collapsed and those from different Pipidae species are in different colors (*X. tropicalis* (XT) -blue, *X. laevis* (XL) - pink and *P. carvalhoi* (PC) - green). Species abbreviations: *Erpetoichthys calabaricus* (EC), *Simochromis diagramma* (SD), *Denticeps clupeoides* (DC) and *Danio rerio* (DR or Dre, according to the Repbase nomenclature of canonical elements). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

divergence reflected in the branch length and k2p distance (Supplementary File 17) combined with the estimative of Ks (Fig. 4) support that the tree inconsistencies could be explained by independent losses in the genomes of related species rather than by HT events. The PC-groups 4 and 5 also seem a result of vertical transmission. However, several other Pipidae groups seem to be involved in horizontal transfer events with fish species. This hypothesis is supported not only by the phylogenetic inconsistencies but also by the very low Ks values found for the TE sequences (Fig. 4).

The Ks value offers a measurement of neutral evolution in the absence of a strong codon bias and consequently, similar values of Ks are expected for TEs and host genes (Silva and Kidwell, 2000). The genes used to test HT present low codon bias with NC varying from 49 to 46 (Supplementary File 18). The Ks values distribution presented by the host genes between fish and Pipidae species is shown in Fig. 4 starting from 0.7 with an average of around 1.4 in all pairwise comparisons. This

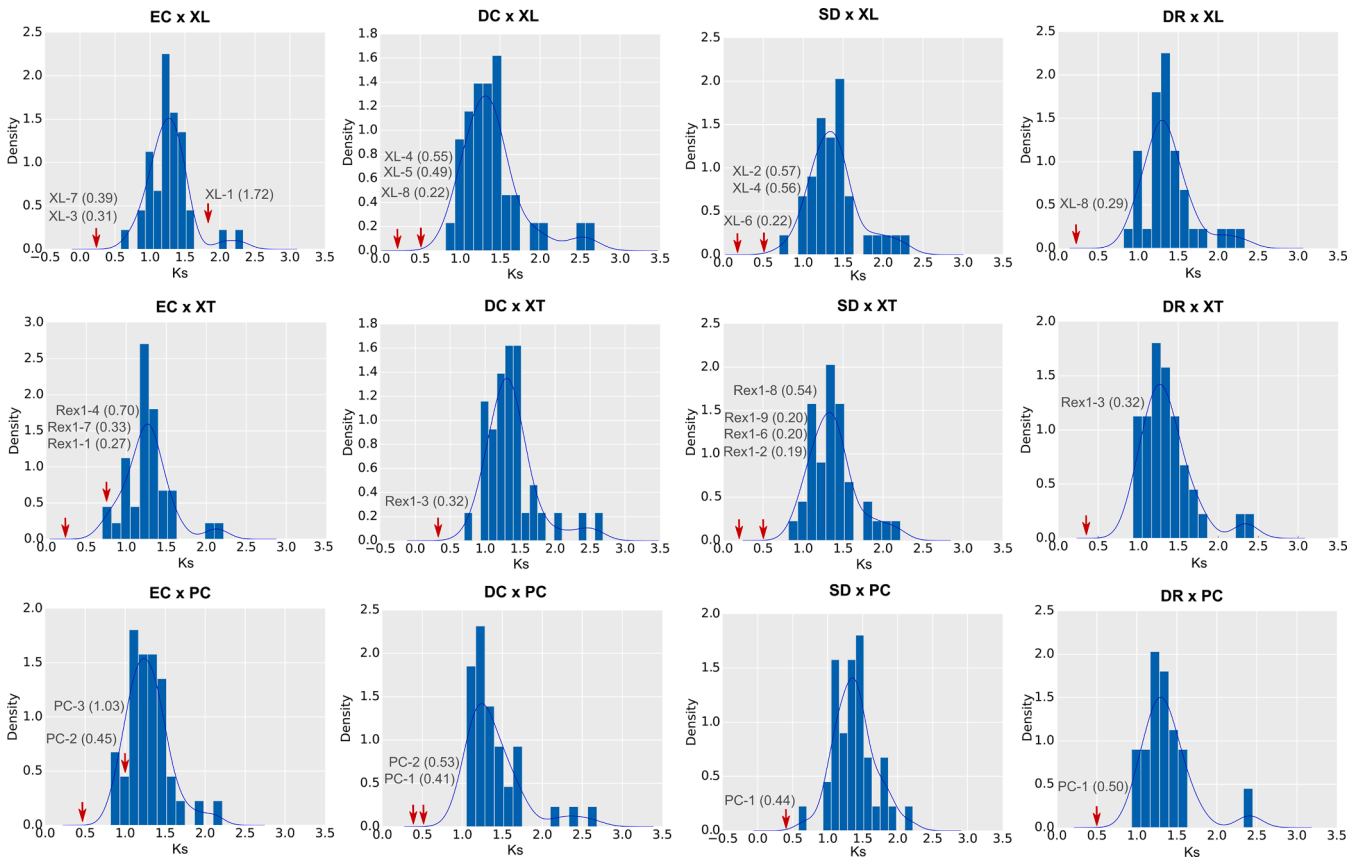

Fig. 4. Histograms of the distribution of the Ks values found for genes between fish and Pipidae species. Red arrows represent the approximate position of the TE Ks values in the graph. The exact value is shown in parenthesis. When more than one species is possibly involved in an HT event, all were included. *XT- X. tropicalis; XL- X. laevis; PC- P. carvalhoi; EC- Erpetoichthys calabaricus, SD- Simochromis diagramma, DC- Denticeps clupeoides and DR- Danio rerio.*

data is in agreement with the Ks average found for the core gene orthologs of vertebrate lineages that diverged 400–500 mya (Zhang et al., 2020) around 1.0 to less than 1.5. Nevertheless, much lower Ks values were found when comparing Pipidae Rex1 sequences with copies from fish species varying from 0.19 to around 0.5 (red arrows in Fig. 4). This high similarity between the Rex1 sequences is also evident in the k2p distribution (Supplementary File 17) where the TE values are found out or in the extreme lower range together with the most conserved genes.

Thus, considering the Ks values found for the TEs, HT is a plausive explanation for several incongruities that we found in the tree. Some HT cases are putative more ancient than others (considering the kS values) and could have involved ancestral species or some other current related species for which the low divergence reverberated in the comparisons we made.

Establishing the direction of the HT is also not a trivial task, thus, for the purpose of description, we hypothesized the possible HT cases considering the current species and avoiding direction determination. Moreover, the possible HT cases were established based on the Ks evidence and considering the sequence tree looking for the most parsimonious scenario and being stringent to count independent transfers and not considering possible HT events among fishes. Following these criteria, we can suggest the following events without disregarding alternative hypotheses:(i) a single event between *P. carvalhoi* (PC-group-1) and a fish; (ii) a single event between *X. laevis* and *D. clupeoides* (XL-group 5); (iii) a single event between *X. tropicalis* and *E. calabaricus* (REX-1_XT); (iv) three single events between *X. tropicalis* and *S. diagramma* (REX1-2_XT, REX1-8_XT and REX1-9_XT); (v) multiples events involving fish and the Pipidae species in the case of the closely related groups REX1-7_XT, XL-groups 2 and 3 and PC-group 2. Although these sequences may have originally been vertically transmitted in Pipidae, multiple HT events would be necessary to describe the disposition of the fish sequences in these groups; (vi) multiple HT events are also likely involving *X. tropicalis* and *X. laevis* with fish species in the case of REX1-6_XT and XL-groups 6 and 7; (vii) at least one HT event between Pipidae and fish species is required to account for the relationship of REX1-3_XT and XL-group 8 with fish species; (viii) and a single HT event could explain the relationship of *X. laevis* from the XL-group 4 and fish species.

4. Discussion

4.1. LINE diversity in Pipidae species

LINE elements are estimated to be as old as eukaryotes, with each the main clade dating back to at least 600 mya (Malik et al., 1999). The Pipidae genomes from South America and Africa species included in our genomic surveys revealed a number of shared LINE clades/superfamilies that probably have been vertically transmitted by the most recent common ancestor of Pipidae and then experienced independent evolutionary ways since the long time of divergence of these species.

We recovered LINE sequences from Rex1, L1, L2 and Tx1 clades in the genome of these three species. CR1 clade was restricted to the African lineages (*X. laevis* and *X. tropicalis*) in the tree since our search strategy was not designed to recover sequences from this clade, however, CR1 sequences can also be found in *P. carvalhoi* (data not shown). On the other hand, RTE clade is shared only between *P. carvalhoi* and *X. laevis* genomes. The clades recovered in our analyses coincide with those described by Hellsten et al. (2010) to *X. tropicalis* genome and most of the recovered copies clearly belong to some family described in the

Repbase. However, for *X. laevis*, several copies retrieved are not related to any described family, indicating that the Repbase TE library for *X. laevis* is highly incomplete and deserves to be updated. In this sense, we reported for the first time in this species, the presence of elements from RTE and Rex1 clades not previously deposited in the Repbase. This work is also the first study of retrotransposons in *P. carvalhoi,* revealing a diversity of LINE elements and showing the importance of sequencing new anuran genomes. Species-specific arrangements of *P. carvalhoi* sequences were prominent on L2 and Rex1 clades, which could indicate recent waves of retrotransposition in this species.

The Pipidae LINEs belong to three groups: CR1, RTE and L1. Rex1, L2 and CR1 clades belong to CR1 group that has been found exclusively in animals (Kojima, 2019). The CR1 clade is widely distributed on tetrapods and their phylogenetic distribution suggests that these elements could be preserved and proliferated during the aquatic to the terrestrial transition of these organisms (Chalopin et al., 2015). L2 is also widely distributed in animals, although discontinuously (Lovšin et al., 2001), while the Rex1 has a more restrict known distribution, being abundant in fish genomes, some amphibians (*X. tropicalis* and *Physalaemus eppifer*) (Hellsten et al., 2010; Nascimento et al., 2015) and also found in cnidarians (Kapitonov et al., 2009) and in the sea urchin *Strongylocentrotus purpuratus* (Rho and Tang, 2009). In *Caiman crocodilus*, Rex1 was identified by fluorescence in situ hybridization but the sequence was not characterized (Oliveira et al., 2021).

The RTE group, represented by the RTE clade in Pipidae, has been found in animals, fungi, plants and algae (Kojima, 2019; Malik and Eickbush, 1998). The absence of this clade in *X. tropicalis* could be explained more parsimoniously by its presence in the ancestor of Pipidae and complete loss in *X. tropicalis*.

Finally, the L1 group is widely distributed in eukaryotes, the clade L1 being found in plants, fungi, green algae and vertebrates while the Tx1 was found in animal species (cnidarian, sea squirt, fish, frog). Hellstein et al (2010) have identified a high number of L1 and Tx1 young families in *X. tropicalis* that are observed by the high number of Repbase families in both clades. Although the genomic searches did not intend to retrieve elements from L1 and Tx1, several Pipidae copies were recovered in these clades close to different *X. tropicalis* families, indicating that *P. carvalhoi* and *X. laevis* also have a high diversity of families that could be further explored. Moreover, the Tx1 families shared among the three Pipidae species reveal an important element to future evaluations. The Tx1-clade are recognized as elements with target site-specificity (Kojima and Fujiwara, 2004), a feature already reported for turtles (Kojima, 2015) and for *X. laevis* (Christensen et al., 2000) and *X. tropicalis* (Hellsten et al., 2010). Elements named as L1-53_XT to L1-56 XT from *X. tropicalis* are preferentially inserted inside of various U2 small RNA genes (Hellsten et al., 2010) such as the elements named Keno (Kojima and Fujiwara, 2004). Copies of Tx1_XT are inserted in the same site in copies of a DNA transposon, the piggyBac-N1_XT. From the few Tx1 copies of *P. carvalhoi* that we were able to analyze the flanking regions, pipa1505315 and pipa445583 were also located close to the U2 spliceosomal RNA sequences. We can suggest that homologous elements from this clade have retained the same genomic pattern of insertion in these species phylogenetically related.

The evolutionary sequence tree recovered here relights the debate on the issue of the nomenclature of TEs and reinforces that the classification should reflect the tree. Despite a number of studies, including data shown here, pointed that Rex elements did not represent a monophyletic group (Burke et al., 2002; Carducci et al., 2018; Lovšin et al., 2001), this nomenclature has contributed to many authors to consider all elements under Rex denomination as a unique group of non-LTR retrotransposons (Carducci et al., 2018; Daniel et al., 2016; Favarato et al., 2017; Nascimento et al., 2015). This misconception could be hiding the real diversity of LINE elements repertoire that compose these genomes.

4.2. Several Rex1 families were likely horizontally transferred between Pipidae and fish species

The Rex1 clade showed a diversity of families shared among Pipidae species besides some recent species-specific waves of retrotransposition (estimates by branches-size), as detected on *P. carvalhoi* genome, for example, a similar evolutionary profile already reported to fish genomes (Chalopin et al., 2015; Coan and Martins, 2018; Volff et al., 2000).

One plausible hypothesis to explain the family diversity currently found in Pipidae genomes could be an ancestral diversity already present in the vertebrate or Pipidae ancestors followed by vertical transmission. However, we observed an unexpected high similarity of some Rex1 families with copies from bony fish species, leading us to employ additional evolutionary analysis, showing also phylogenetic inconsistencies on the nucleotide sequence tree inferred to Rex1 copies from Pipidae and fish species, in which we detected unusual close relationships among elements from amphibians and fishes.

The ancient evolutionary origin of these families and independent events of loss among lineages could generate similar signatures to those we found in the tree. This reasoning is sufficient to explain some cases in which elements from Pipidae and fishes species exhibit a high level of sequence divergence. However, we recovered several cases in which this explanation is not sufficient to justify the high similarity between TE sequences from distantly related species. HT transfer was tested here and it is strongly suggested in several cases. HT has been evocated to explain several incongruences detect on evolutionary sequences trees of TEs (to review de Melo et al., 2020; Wallau et al., 2018) and recent new approaches have shown the more frequent occurrence of these events than previously predict (de Melo et al., 2020; Zhang et al., 2020). Recently, Zhang et al (2020) reported an expressive high number of HT events between vertebrates, 175 involving Rex1 elements called Rex-Babar, most among teleost fishes. We also observed in our data that some Ks values among the four fish species suggest HT. Although it was not discussed in the main text, these last authors also found some HT events of Rex1 elements between fishes and amphibians (nine cases involving *X. tropicalis* and *X. laevis* and one involving *N. parkeri*) (Zhang et al 2020). Thus, the data provided here and by other authors indicate a high rate of HT events of Rex1 families between fish and Pipidae.

Aquatic habitats have been suggested to be more likely to exchange TEs than terrestrial ones (Metzger et al., 2018; Wang and Liu, 2016) and HT events between aquatic lifestyle species are being continuously reported such as the case of an LTR-retrotransposon, initially isolated from shell mollusks and analyzed in many other bivalve species, finding evidence of HT (Metzger et al., 2018). Another study identified several TEs, mostly LINEs, in the Pacific white shrimp (*Litopenaeus vannamei*) which seem to be derived from HT from other aquatic organisms (Wang and Liu, 2016). Galbraith et al. (2020) have reported HT cases of six LINE retrotransposons in the marine snake (*Aipysurus laevis*) that were probably acquired horizontally from fish or marine invertebrate parasites. The aquatic environment could be favorable to HT, due to the ability of particles to spread without exposure to ultraviolet rays or dry air from the terrestrial environment (Metzger et al., 2018). On the other hand, Zhang et al. (2020) that have detected a high number of HT among vertebrate species, have tested whether aquatic vertebrates were more feasible to transfer TEs than terrestrial ones and found no strong evidence to this hypothesis suggesting further tests are needed, with a larger sampling of genomes. Moreover, HT events appear not to be restricted to Pipidae lineages and a similar pattern was already reported to the terrestrial frog *Oophaga pumilio* in which multiple HT events of TEs, mainly DNA transposons, have also been suggested involving fish (Rogers et al., 2018).

Undoubtedly, ecological connections such as predation, symbiosis and parasitism are necessary for the occurency of HT and these interactions can be facilitated among species that occupy the same niche (Gilbert and Feschotte, 2018; Venner et al., 2017; Wallau et al., 2012). Also, certain classes of organisms displaying pervasive presence in many

ecosystems could have a possible role as reservoirs of TEs that can be further redistributed among other species (Palazzo et al., 2021).

In recent years, some studies are showing the capacity of extracellular vesicles (EVs) in mediating HT transfer of genes or TEs in cell culture (Fischer et al., 2016; Kawamura et al., 2019). It is known that the production of EVs is a common cellular feature to the three domains of life (Deatherage et al., 2012) and a diverse pool of DNA from several phyla was recovered from vesicles isolated from marine ecosystems (Biller et al., 2014). Although the majority of those sequences were from bacteria, EVs were detected in various biological fluids in multicellular eukaryotes (Colombo et al., 2014). The EVs are being considered important players in cell-to-cell communication in aquatic environments, where the effector molecules are protected from the harmful activity of nucleases and proteases, and they serve as vectors for horizontal gene transfer among prokaryotes and viruses (Schatz and Vardi, 2018). While the biological relevance of EVs in the HT process between multicellular eukaryotes is unknown, mainly in aquatic environments, it is an attractive hypothesis. In this sense, we could hypothesize that EVs containing Rex RNAs could be involved in the process of HTs detected here.

5. Conclusions

In conclusion, the data presented in this work indicate that the elements LINEs recovered for anurans are ancient, probably already present in the ancestor of fish and frogs and represent an important element in the evolution of the genomes in this group, and waves of retrotransposition in specific strains may have occurred independently during the diversification of species. The Rex1 clade showed a great diversity of families, some of them with high similarity to Rex1 copies from bony fish species, suggesting horizontal transfer events.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) [07717-6, 2016], Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Paraná [51370.545.47347.19112018], Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/PROAP – Finance Code 001) for the scholarships provided to JMG and CG.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107393.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bartolomé, C., Bello, X., Maside, X., 2009. Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. Genome Biol. 10, 1–11. https://doi.org/10.1186/gb-2009-10-2-r22.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Crollius, H.R., Guiguen, Y., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5, 3657. https://doi.org/10.1038/ncomms4657.

Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., Ortí, G., 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17, 162. https://doi.org/10.1186/s12862-017-0958-3.

Biémont, C., Vieira, C., 2006. Junk DNA as an evolutionary force. Nature 443, 521–524. https://doi.org/10.1038/443521a.

Biller, S.J., Schubotz, F., Roggensack, S.E., Thompson, A.W., Summons, R.E., Chisholm, S.W., 2014. Bacterial vesicles in marine ecosystems. Science (80-.) 343, 183–186. https://doi.org/10.1126/science.1243457.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Burke, W.D., Malik, H.S., Rich, S.M., Eickbush, T.H., 2002. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, Giardia lamblia. Mol. Biol. Evol. 19, 619–630. https://doi.org/10.1093/oxfordjournals.molbev.a004121.

Cannatella, D., 2015. Xenopus in Space and Time: Fossils, Node Calibrations, Tip-Dating, and Paleobiogeography. Cytogenet. Genome Res. 145 (3-4), 283–301. https://doi.org/10.1159/000438910.

Capella-gutiérrez, S., Silla-martínez, J.M., Gabaldón, T., 2009. trimAl : a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinforma. Appl. note 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carducci, F., Barucca, M., Canapa, A., Biscotti, M.A., 2018. Rex Retroelements and Teleost Teleost Genomes : An Overview. Int. J. Mol. Sci. 19, 1–15. https://doi.org/10.3390/ijms19113653.

Chalopin, D., Niville, M., Plard, F., Galiana, D., Volff, J., 2015. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. Genome Biol. Evol. 7, 567–580. https://doi.org/10.1093/gbe/evv005.

Christensen, S., Pont-Kingdon, Geneviève, Carroll, D., 2000. Target Specificity of the Endonuclease from the Xenopus laevis Non-Long Terminal Repeat Retrotransposon, Tx1L. Mol. Cell. Biol. 20 (4), 1219–1226.

Coan, R.L.B., Martins, C., 2018. Landscape of transposable elements focusing on the B chromosome of the cichlid fish astatotilapia latifasciata. Genes (Basel). 9, 1–18. https://doi.org/10.3390/genes9060269.

Colombo, M., Raposo, G., Théry, C., 2014. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. Annu. Rev. Cell Dev. Biol. 30 (1), 255–289. https://doi.org/10.1146/annurev-cellbio-101512-122326.

Daniel, S.N., Penitente, M., Silva, D.M.Z.A., Hashimoto, D.T., Ferreira, D.C., Foresti, F., Porto-Foresti, F., 2016. Organization and Chromosomal Distribution of Histone Genes and Transposable Rex Elements in the Genome of Astyanax bockmanni (Teleostei, Characiformes). Cytogenet. Genome Res. 146 (4), 311–318. https://doi.org/10.1159/000441613.

Melo, E.S.d., Wallau, G.L., Malik, H.S., 2020. Mosquito genomes are frequently invaded by transposable elements through horizontal transfer. PLoS Genet. 16 (11), e1008946. https://doi.org/10.1371/journal.pgen.1008946.

Deatherage, B.L., Cookson, B.T., Andrews-Polymenis, H.L., 2012. Membrane vesicle release in bacteria, eukaryotes, and archaea: A conserved yet underappreciated aspect of microbial life. Infect. Immun. 80 (6), 1948–1957. https://doi.org/10.1128/IAI.06014-11.

Edwards, R.J., Tuipulotu, D.E., Amos, T.G., Meally, D.O., Richardson, M.F., Russell, T.L., Vallinoto, M., Carneiro, M., Ferrand, N., Wilkins, M.R., Sequeira, F., Rollins, L.A., Holmes, E.C., Shine, R., White, P.A., 2018. Draft genome assembly of the invasive cane toad, Rhinella marina. GigaScience 7, 1–13. https://doi.org/10.1093/gigascience/giy095.

Eickbush, T.H., Malik, H.S., 2002. Origins and Evolution of Retrotransposons. In: Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. American Society of Microbiology Press, Washington, pp. 1111–1144. https://doi.org/10.1128/9781555817954.ch49.

Elliott, T.A., Gregory, T.R., 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos. Trans. R. Soc. B Biol. Sci. 370, 1–10. https://doi.org/10.1098/rstb.2014.0331.

Farré, M., Bosch, M., López-Giráldez, F., Ponsà, M., Ruiz-Herrera, A., Liberles, D., 2011. Assessing the Role of Tandem Repeats in Shaping the Genomic Architecture of Great Apes. PLoS ONE 6 (11), e27239. https://doi.org/10.1371/journal.pone.0027239.

Favarato, R.M., Ribeiro, L.B., Feldberg, E., Matoso, D.A., 2017. Chromosomal mapping of transposable elements of the rex family in the bristlenose catfish, ancistrus (siluriformes, loricariidae), from the amazonian region. J. Hered. 108, 254–261. https://doi.org/10.1093/jhered/esw084.

Feschotte, C., Pritham, E.J., 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. Annu. Rev. Genet. 41 (1), 331–368. https://doi.org/10.1146/annurev.genet.40.110405.090448.

Fischer, S., Cornils, K., Speiseder, T., Badbaran, A., Reimer, R., Indenbirken, D., Grundhoff, A., Brunswig-Spickenheier, B., Alawi, M., Lange, C., Camussi, G., 2016. Indication of horizontal DNA gene transfer by extracellular vesicles. PLoS ONE 11 (9), e0163665. https://doi.org/10.1371/journal.pone.0163665.

Galbraith, J.D., Ludington, A.J., Suh, A., Sanders, K.L., Adelson, D.L., 2020. New environment, new invaders - Repeated horizontal transfer of lines to sea snakes. Genome Biol. Evol. 12, 2370–2383. https://doi.org/10.1093/GBE/EVAA208.

Gilbert, C., Feschotte, C., 2018. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr. Opin. Genet. Dev. 49, 15–24. https://doi.org/10.1016/j.gde.2018.02.007.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., Kumar, S., 2015. Tree of Life Reveals Clock-Like Speciation and Diversification. Mol. Biol. Evol. 32, 835–845. https://doi.org/10.1093/molbev/msv037.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., Blitz, I.L., Blumberg, B., Dichmann, D.S., Dubchak, I., Amaya, E., Detter, J.C., Fletcher, R., Gerhard, D.S., Goodstein, D., Graves, T., Grigoriev, I.V., Grimwood, J., Kawashima, T., Lindquist, E., Lucas, S.M., Mead, P.E., Mitros, T., Ogino, H., Ohta, Y., Poliakov, A.V., Pollet, N., Robert, J., Salamov, A., Sater, A.K., Schmutz, J., Terry, A., Vize, P.D., Warren, W.C., Wells, D., Wills, A., Wilson, R.K., Zimmerman, L.B., Zorn, A.M., Grainger, R., Grammer, T., Khokha, M.K., Richardson, P.M., Rokhsar, D.S., 2010. The Genome of the western Clawed Frog Xenopus tropicalis. Science (80-.) 328 (5978), 633–636.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 9 (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

Irisarri, I., Vences, M., Mauro, D.S., Glaw, F., Zardoya, R., 2011. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. BMC Evol. Biol. 11, 1–10. https://doi.org/10.1186/1471-2148-11-114.

Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V., 2007. Repetitive Sequences in Complex Genomes : Structure and Evolution. Annu. Rev. Genomics Hum. Genet. 8, 241–259. https://doi.org/10.1146/annurev.genom.8.080706.092416.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467. https://doi.org/10.1159/000084979.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M., Celniker, S.E., 2002. The transposable elements of the Drosophila melanogaster euchromatin : a genomics perspective. Genome Biol. 3, 1–20. https://doi.org/10.1186/gb-2002-3-12-research0084.

Kapitonov, V.V., Tempel, S., Jurka, J., 2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene 448 (2), 207–213. https://doi.org/10.1016/j.gene.2009.07.019.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service : multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 1–7 https://doi.org/10.1093/bib/bbx108.

Kawamura, Y., Sanchez Calle, A., Yamamoto, Y., Sato, T.A., Ochiya, T., 2019. Extracellular vesicles mediate the horizontal transfer of an active LINE-1 retrotransposon. J. Extracell. Vesicles 8, 1–17. https://doi.org/10.1080/20013078.2019.1643214.

Keinath, M.C., Timoshevskiy, V.A., Timoshevskaya, N.Y., Tsonis, P.A., Voss, S.R., Smith, J.J., 2015. Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. Sci. Rep. 5, 1–13. https://doi.org/10.1038/srep16413.

Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115, 49–63. https://doi.org/10.1023/A:1016072014259.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (2), 111–120. https://doi.org/10.1007/BF01731581.

Kojima, K.K., 2019. Structural and sequence diversity of eukaryotic transposable elements. Genes Genet. Syst. 94 (6), 233–252. https://doi.org/10.1266/ggs.18-00024.

Kojima, K.K., 2015. A new class of SINEs with snRNA gene-derived heads. Genome Biol. Evol. 7, 1702–1712. https://doi.org/10.1093/gbe/evv100.

Kojima, K.K., Fujiwara, H., 2004. Cross-Genome Screening of Novel Sequence-Specific Non-LTR Retrotransposons: Various Multicopy RNA Genes and Microsatellites Are Selected as Targets. Mol. Biol. Evol. 21, 207–217. https://doi.org/10.1093/molbev/msg235.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., Gascuel, O., 2019. NGPhylogeny.fr: New generation phylogenetic services for non-specialists. Nucleic Acids Res. 47, W260–W265. https://doi.org/10.1093/nar/gkz303.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and. Nucleic Acids Res. 47, 256–259. https://doi.org/10.1093/nar/gkz239.

Li, W.-H., 1993. Unbiased Estimation of the Rates of Synonymous and Nonsynonymous Substitution. J. Mol. Evol. 36 (1), 96–99.

López-Flores, I., Garrido-Ramos, M., 2012. The Repetitive DNA Content of Eukaryotic Genomes. In: Garrido-Ramos, M. (Ed.), Repetitive DNA. Karger, Genome Dynamic, pp. 1–28.

Lovšin, N., Gubenšek, F., Kordiš, D., 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in deuterostomia. Mol. Biol. Evol. 18, 2213–2224. https://doi.org/10.1093/oxfordjournals.molbev.a003768.

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res. 48, D265–D268. https://doi.org/10.1093/nar/gkz991.

Ludwig, A., Valente, V.L.d.S., Loreto, E.L.S., 2008. Multiple invasions of Errantivirus in the genus Drosophila. Insect Mol. Biol. 17 (2), 113–124. https://doi.org/10.1111/j.1365-2583.2007.00787.x.

Malik, H.S., Burke, W.D., Eickbush, T.H., 1999. The Age and Evolution of Non-LTR Retrotransposable Elements. Mol. Biol. Evol. 16, 793–805. https://doi.org/10.1093/oxfordjournals.molbev.a026164.

Malik, H.S., Eickbush, T.H., 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol. Biol. Evol. 15 (9), 1123–1134. https://doi.org/10.1093/oxfordjournals.molbev.a026020.

Marchler-bauer, A., Bryant, S.H., 2004. CD-Search : protein domain annotations on the fly. Nucleic Acids Res. 32, 327–331. https://doi.org/10.1093/nar/gkh454.

Metzger, M.J., Paynter, A.N., Siddall, M.E., Goff, S.P., 2018. Horizontal transfer of retrotransposons between bivalves and other aquatic species of multiple phyla. Proc. Natl. Acad. Sci. U. S. A. 115 (18), E4227–E4235. https://doi.org/10.1073/pnas.1717227115.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In: Gateway Computing Environments Workshop (GCE), New Orleans, LA, pp. 1–8.

Nascimento, J., Baldo, D., Lourenço, L.B., 2015. First insights on the retroelement Rex1 in the cytogenetics of frogs. Mol. Cytogenet. 8, 1–9. https://doi.org/10.1186/s13039-015-0189-5.

Nicholas, K.B., Nicholas, H.B.J., 1997. Genedoc: a tool for editing and annotating multiple sequence alignments. Distrib. by authors.

Noronha, R.C.R., Barros, L.M.R., Araújo, R.E.F., Marques, D.F., Nagamachi, C.Y., Martins, C., Pieczarka, J.C., 2016. New insights of karyoevolution in the Amazonian turtles Podocnemis expansa and Podocnemis unifilis (Testudines, Podocnemidae). Mol. Cytogenet. 9, 1–9. https://doi.org/10.1186/s13039-016-0281-5.

Oliveira, V.C.S., Viana, P.F., Gross, M.C., Feldberg, E., Da Silveira, R., de Bello Cioffi, M., Bertollo, L.A.C., Schneider, C.H., 2021. Looking for genetic effects of polluted anthropized environments on Caiman crocodilus crocodilus (Reptilia, Crocodylia): A comparative genotoxic and chromosomal analysis. Ecotoxicol. Environ. Saf. 209, 111835. https://doi.org/10.1016/j.ecoenv.2020.111835.

Palazzo, A., Escuder, E., D'Addabbo, P., Lovero, D., Marsano, R.M., 2021. A genomic survey of Tc1-mariner transposons in nematodes suggests extensive horizontal transposon transfer events. Mol. Phylogenet. Evol. 158, 107090. https://doi.org/10.1016/j.ympev.2021.107090.

Pamilo, P., Bianchi, N.O., 1993. Evolution of the Z fx and Zfy Genes : Rates and Interdependence between the Genes. Mol. Biol. Evol. 10, 271–281.

Pyron, R.A., 2010. A Likelihood Method for Assessing Molecular Divergence Time Estimates and the Placement of Fossil Calibrations. Syst. Biol. 59, 185–194. https://doi.org/10.1093/sysbio/syp090.

Rho, M., Tang, H., 2009. MGEScan-non-LTR: Computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. Nucleic Acids Res. 37 (21), e143. https://doi.org/10.1093/nar/gkp752.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16 (6), 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Rogers, R.L., Zhou, L., Chu, C., Marquez, R., Corl, A., Linderoth, T., Freeborn, L., MacManes, M.D., Xiong, Z., Zheng, J., Guo, C., Xun, X., Kronforst, M.R., Summers, K., Wu, Y., Yang, H., Richards-Zawacki, C.L., Zhang, G., Nielsen, R., 2018. Genomic takeover by transposable elements in the strawberry poison frog. Mol. Biol. Evol. 35, 2913–2927. https://doi.org/10.1093/molbev/msy185.

Safian, D., Wiegertjes, G.F., Pollux, B.J.A., 2021. The Fish Family Poeciliidae as a Model to Study the Evolution and Diversification of Regenerative Capacity in Vertebrates. Front. Ecol. Evol. 9, 167. https://doi.org/10.3389/fevo.2021.613157.

Saint-Leandre, B., Nguyen, S.C., Levine, M.T., 2019. Diversification and collapse of a telomere elongation mechanism. Genome Res. 29 (6), 920–931. https://doi.org/10.1101/gr.245001.118.

Schatz, D., Vardi, A., 2018. Extracellular vesicles — new players in cell–cell communication in aquatic environments. Curr. Opin. Microbiol. 43, 148–154. https://doi.org/10.1016/j.mib.2018.01.014.

Seppey, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.

Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., van Heeringen, S.J., Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J.B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S.S., van Kruijsbergen, I., Shu, S., Carlson, J., Kinoshita, T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., Mitros, T., Mozaffari, S.V., Suzuki, Y., Haramoto, Y., Yamamoto, T.S., Takagi, C., Heald, R., Miller, K., Haudenschild, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D.S., Flajnik, M.F., Houston, D.W., Shendure, J., DuPasquier, L., Vize, P.D., Zorn, A.M., Ito, M., Marcotte, E.M., Wallingford, J.B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G.J.C., Fujiyama, A., Harland, R.M., Taira, M., Rokhsar, D.S., 2016. Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538, 336–343. https://doi.org/10.1038/nature19840.

Silva, J.C., Kidwell, M.G., 2000. Horizontal transfer and selection in the evolution of P elements. Mol. Biol. Evol. 17, 1542–1557. https://doi.org/10.1093/oxfordjournals.molbev.a026253.

Sotero-Caio, C.G., Platt II, R.N., Suh, A., Ray, D.A., 2017. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. Genome Biol. 9, 161–177. https://doi.org/10.1093/gbe/evw264.

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., Zhong, L.i., Wang, L.u., Wu, D.-D., Zhang, B.-L., Zhu, C.-L., Yang, M.-M., Chen, H.-M., Li, F., Zhou, L., Feng, S.-H., Huang, C., Zhang, G.-J., Irwin, D., Hillis, D.M., Murphy, R.W., Yang, H.-M., Che, J., Wang, J., Zhang, Y.-P., 2015. Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes.

Proc. Natl. Acad. Sci. U. S. A. 112 (11), E1257–E1262. https://doi.org/10.1073/pnas.1501764112.
Venner, S., Miele, V., Terzian, C., Biémont, C., Daubin, V., Feschotte, C., Pontier, D., 2017. Ecological networks to unravel the routes to horizontal transposon transfers. PLoS Biol. 15 (2), e2001536. https://doi.org/10.1371/journal.pbio.2001536.
Volff, J.-N., Körting, C., Froschauer, A., Sweeney, K., Schartl, M., 2001. Non-LTR Retrotransposons Encoding a Restriction Enzyme-Like Endonuclease in Vertebrates. Mol. Biol. Evol. 52 (4), 351–360. https://doi.org/10.1007/s002390010165.
Volff, J., Korting, C., Schartl, M., 2000. Multiple Lineages of the Non-LTR Retrotransposon Rex1 with Varying Success in Invading Fish Genomes. Mol. Biol. Evol. 17, 1673–1684.
Volff, J.N., Korting, C., Sweeney, K., Schartl, M., 1999. The Non-LTR Retrotransposon Rex3 from the Fish Xiphophorus is Widespread Among Teleosts. Mol. Biol. Evol. 16 (11), 1427–1438.
Volff, J.N., Lehrach, H., Reinhardt, R., Chourrout, D., 2004. Retroelement dynamics and a novel type of chordate retrovirus-like element in the miniature genome of the tunicate Oikopleura dioica. Mol. Biol. Evol. 21, 2022–2033. https://doi.org/10.1093/molbev/msh207.
Wallau, G.L., Ortiz, M.F., Loreto, E.L.S., 2012. Horizontal Transposon Transfer in Eukarya: Detection, Bias, and Perspectives. Genome Biol. Evol. 4, 801–811. https://doi.org/10.1093/gbe/evs055.
Wallau, G.L., Vieira, C., Loreto, É.L.S., 2018. Genetic exchange in eukaryotes through horizontal transfer: Connected by the mobilome. Mob. DNA 9, 1–16. https://doi.org/10.1186/s13100-018-0112-9.
Wang, X., Liu, X., 2016. Close ecological relationship among species facilitated horizontal transfer of retrotransposons. BMC Evol. Biol. 16, 1–13. https://doi.org/10.1186/s12862-016-0767-0.
Wells, J.N., Feschotte, C., 2020. A Field Guide to Eukaryotic Transposable Elements. Annu. Rev. Genet. 54, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145.
Wicker, T., Sabot, F., Hua-van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., Sanmiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat. Genet. 8, 973–982. https://doi.org/10.1038/nrg2165.
Xia, X., 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Mol. Biol. Evol. 35, 1550–1552. https://doi.org/10.1093/molbev/msy073.
Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18 (5), 821–829. https://doi.org/10.1101/gr.074492.107.
Zhang, H.H., Peccoud, J., Xu, M.R.X., Zhang, X.G., Gilbert, C., 2020. Horizontal transfer and evolution of transposable elements in vertebrates. Nat. Commun. 11, 1–10. https://doi.org/10.1038/s41467-020-15149-4.