Title: Evolutionary dynamics of *DIRS-like* and *Ngaro-like* retrotransposons in *Xenopus laevis* and *Xenopus tropicalis* genomes.

Authors: Camilla Borges Gazolla^{†*}; Adriana Ludwig^{‡*}; Joana de Moura Gama[§]; Daniel Pacheco Bruschi[†]

[†] Laboratório de Citogenética evolutiva e Conservação Animal (LabCeca), Departamento de Genética, Universidade Federal do Paraná

[‡] Laboratório de Ciências e Tecnologias Aplicadas em Saúde (LaCTAS), Instituto Carlos Chagas – Fiocruz-PR, Curitiba, PR, Brazil.

[§] Departamento de Biologia Estrutural and Funcional, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária, 13083-863, Campinas, SP, Brazil

* Equal contributions

Orcid: Camilla Borges Gazolla – 0000-0003-2520-8038

Joana de Moura Gama - 0000-0003-0378-6670

Adriana Ludwig - 0000-0001-9170-7156

Daniel Pacheco Bruschi - 0000-0003-2181-1788

Title: Differential evolutionary dynamics of DIRS-like and Ngaro-like retrotransposons in *Xenopus laevis* and *Xenopus tropicalis* genomes.

Running title: Retrotransposons DIRS in *Xenopus*

Keywords

Anura; transposable elements; DIRS; retrotransposon

Author corresponding: Daniel Pacheco Bruschi

Address: Centro Politécnico, Av. Cel. Francisco H. dos Santos, 100 - Jardim das Américas, Curitiba - PR, 81531-980

Phone number: +55 41 9216-2802

Email: danielpachecobruschi@gmail.com/ danielbruschi@ufpr.br

ABSTRACT

Anuran genomes have a large number and diversity of transposable elements, but are little explored, mainly in relation to their molecular structure and evolutionary dynamics. Here, we investigated the retrotransposons containing tyrosine recombinase (YR) (order DIRS) in the genome of *Xenopus tropicalis* and *Xenopus laevis*. These anurans show $2n = 20$ and the $2n = 36$ karyotypes, respectively. They diverged about 48 million years ago (mya) and *X. laevis* had an allotetraploid origin (around 17-18 mya). Our investigation is based on the analysis of the molecular structure and the phylogenetic relationships of 95 DIRS families of *Xenopus* belonging to *DIRS-like* and *Ngaro-like* superfamilies. We were able to identify molecular signatures in the 5' and 3' non-coding terminal regions, preserved open reading frames (ORFs) and conserved domains that are specific to distinguish each superfamily. We recognize two ancient amplification waves of *DIRS-like* elements that occurred in the ancestor of both species and a higher density of the old/degenerate copies detected in the *X. laevis*. *X. tropicalis* showed more recent amplification waves estimated around 16 mya and 3.2 mya and corroborate with high diversity-level of families in this species and with transcriptional activity evidence. *Ngaro-like* elements presented less diversity and quantity in the genomes, although potentially active copies were also found. Our findings highlight a differential diversity-level and evolutionary dynamics of the YR retrotransposons in the diploid *X. tropicalis* and *X. laevis* species expanding our comprehension of the behavior of these elements in both genomes during the diversification process.

INTRODUCTION

Transposable elements (TEs) are the most variable feature of the vertebrate genome, and their role in shaping genomic diversity has attracted considerable interest in recent years (Bourque et al. 2018; Wicker et al. 2018). An exceptional diversity of TEs has been reported in all the amphibian genomes sequenced so far (Hellsten et al. 2010; Sun et al. 2012; Jiang et al. 2015; Session et al. 2016; Hammond et al. 2017; Edwards et al. 2018; Seidl et al. 2019). In *Xenopus tropicalis*, a model organism for genomic studies, TEs represent approximately one-third of the genome (Hellsten et al. 2010). Despite the considerable abundance of TEs in genome annotation, the diversity, molecular structure, and evolutionary dynamics of these elements are still poorly understood. The DIRS elements are a good example of this richness that has not been explored.

Retrotransposons of the order DIRS are widely distributed in the eukaryote genomes (Wicker et al. 2007), except for the birds and mammals (Poulter and Butler 2015). The unifying feature of these elements is that they encode a tyrosine recombinase (YR) which participates in the process of integrating the element into the genome (Poulter and Butler 2015). Other retrotransposons employ endonucleases (LINEs and PLEs) or DDE-type integrase (LTRs) (Wicker et al. 2007).

The DIRS elements were named in recognition of the first retrotransposon containing YR to be described, *DIRS-1*, which was found in the slime mold *Dictyostelium discoideum* (Cappello et al. 1985). This order can be divided into four superfamilies based on sequence structure and phylogeny: *DIRS-like*, *Ngaro-like*, *PAT-*

like, and *VIPER-like* (Ribeiro et al. 2019). In general, the *DIRS* elements have three open reading frames (ORFs). The first ORF corresponds to a gag-like domain, the second corresponds to the reverse transcriptase (RT) and RNase H (RH), and the third corresponds to the YR. Another characteristic of these elements is that the ORFs frequently overlap and have terminal repeats that vary in structure among the superfamilies (Poulter and Goodwin 2005; Ribeiro et al. 2019).

The *DIRS-like* elements can present a conserved methyltransferase (MT) domain downstream from the RT/RH, although the function of this domain is still unknown (Goodwin et al. 2004; Poulter and Butler 2015). The non-coding portion varies in its sequence among the elements, although its basic structure is composed of inverted terminal repeats (ITRs) and an internal complementary region (ICR), which is complementary to the beginning of ITR5' and the end of the ITR3' (Cappello et al. 1985; Poulter and Butler 2015).

The *Ngaro-like* elements were described after the *DIRS-like* and are distinguished by their split direct repeats (SDR), composed of A1 in the 5' end and B1, A2, and B2 in the 3' end, where A1 and A2 are identical, as are B1 and B2 (Goodwin et al. 2004). These elements do not contain the MT-like domain found in the *DIRS-like*, although in amphibians, they contain an ORF encoding a hydrolase domain (Hydro - SGNH) after the YR, but with no proven function (Goodwin and Poulter 2004; Poulter and Butler 2015).

The *PAT-like* elements are phylogenetically closely related to the *DIRS-like* elements (Goodwin and Poulter 2001; Goodwin and Poulter 2004; Poulter and Goodwin 2005), although these two groups are not always monophyletic (Ribeiro et al. 2019), and they can be differentiated by structural differences in the terminal repeats. Such as for *Ngaro-like*, *PAT-like* elements are composed of SDRs (Poulter and Goodwin 2005;

Ribeiro et al. 2019). The *VIPER-like* elements also have SDRs, and form a distinct group of retrotransposons restricted to the protozoans of the order Kinetoplastida (Ribeiro et al. 2019).

In the Anura, both *DIRS-like* and *Ngaro-like* elements have been described in *X. tropicalis* and *X. laevis* (Goodwin et al. 2004; Hellsten et al. 2010; Poulter and Butler 2015). These species are found across sub-Saharan Africa and have an aquatic life that distinguishes them from other anurans (Hellsten et al. 2010). The *X. tropicalis* karyotype is composed of $2n=20$ chromosomes with an estimated genome size of 1.7 Gbp (Hellsten et al. 2010) while the *X. laevis* karyotype has a diploid number of $2n=36$ chromosomes, which originated from a process of allopolyploidy, with an estimated size of 3.1 Gbp (Session et al. 2016). The available estimates indicate that 1% of the *X. tropicalis* genome is composed of distinct families of DIRS, some of which may still be active (Hellsten et al. 2010). Evidence of the transcriptional activity of the DIRS elements has already been found in both species (Poulter and Butler 2015), which highlights the possible role of these elements in genome function and evolution.

In the present study, we evaluated the diversity, molecular structure and evolutionary dynamics of the elements of the order DIRS in *X. tropicalis* and *X. laevis*. We identified the structural characteristics of the YR retrotransposons of the DIRS order in both genomes and described diagnostic characteristics for the best differentiation of the elements of the *DIRS-like* and *Ngaro-like* superfamilies and evaluated the evolutionary dynamics of these superfamilies in these genomes.

MATERIAL AND METHODS

An extract containing all the elements identified as DIRS was obtained from the Rebbase database (Jurka 2000) version 23.11. All the sequences from *X. tropicalis* and *X. laevis* were selected and analyzed using the NCBI “Open Reading Frame Finder” (ORFfinder) (<https://www.ncbi.nlm.nih.gov/orffinder/>) to identify ORFs with default parameters (“minimal ORF length (nt)” = 75; “Genetic code”: 1. Standard; “ORF start codon to use”: ATG only). The presence of conserved domains was analyzed using the NCBI “Conserved Domains Search Service” (CD-Search) (Marchler-Bauer and Bryant 2004) with an e-value threshold adjusted to 0.1. The presence of inverted terminal repeats (ITRs), internal complementary region (ICRs), and split direct repeats (SDRs) was investigated using NCBI BLASTn with the same sequence as query and subject, selecting the options “Align two or more sequences” and “somewhat similar sequences (blastn)”, with the “word size” parameter being adjusted to the minimum and the e-value threshold was 10.

Three families of *X. tropicalis* were selected for the analysis of copies in the genome, including one *DIRS-like* family (DIRS-37_XT) and two *Ngaro-like* families (DIRS-53_XT and DIRS-54_XT). We chose the DIRS-37_XT and DIRS-53_XT families as queries because they have the conserved structure of the ORFs and the complete domains, as well as the characteristic repeats for each superfamily. The DIRS-54_XT family was also used as a query to expand the searches of *Ngaro-like* even despite not having conserved terminal repeats.

The amino acid sequences corresponding to the RTs of both elements were used as queries in online tBLASTn searches against the *X. tropicalis* (GCA_000004195.4)

and *X. laevis* (GCA_001663975.1) genomes. The first ten hits were retrieved with 3 kb of both upstream and downstream regions. All the copies retrieved were analyzed for the identification of the ORFs, the conserved domains, and the repetitive regions as described above.

The evolutionary analyses were based on the alignment of the RT amino acid sequences, including the following sequences: (i) the consensus sequences of the DIRS families of *X. laevis* and *X. tropicalis* recovered from Rebbase; (ii) the copies that are homologous to the DIRS-37_XT, DIRS-53_XT, and DIRS-54_XT families retrieved from the *X. tropicalis* and *X. laevis* genomes; and (iii) elements known to belong to the different superfamilies of the order DIRS, *Ngaro-like* - Ngaro1_DR (AY152729 - *Danio rerio*) and Lv_Ngaro2 (AGCV01398517 - *Lytechinus variegatus*), *PAT-like* - SkowPAT (Rebbase - *Saccoglossus kowalevskii*), and *PAT* (Q26106 - *Panagrellus redivivus*), and *DIRS-like* - DIRS-1_Acar (Rebbase - *Anolis carolinensis*) and DIRS-5_CBP (Rebbase - *Chrysemys picta bellii*).

The sequences were aligned using the PSI-coffee tool (Notredame et al. 2000), with Genedoc 2.7 (Nicholas *et al.*, 1997) being used for sequence manipulation and editing. Most of the RT sequences contained around 200 amino acids (aa), and sequences with less than 70% coverage were excluded from the matrix. The MegaX program (Kumar et al. 2018) was used to determine the best amino acid substitution model. The phylogenetic tree was reconstructed using Bayesian inference (BI), run in MrBayes 3.2.6 (Ronquist et al. 2012) based on the LG+G model. The Markov Chain Monte Carlo (MCMC) was run for 10,000,000 generations, sampled every 1000 generations, with 25% of the initial results being discarded as burn-in. The final tree was visualized and edited using iTOL (Letunic and Bork 2019).

The copies were named *a priori* according to the family used as the query, abbreviated to D37, D53 and D54, followed by the number of the copy referring to the order in which the sequence was recovered, while “XT” and “XL” are acronyms for *X. tropicalis* and *X. laevis*, respectively.

For the landscape analysis, the consensus sequences available in the Repbase for each *DIRS-like* family of *X. tropicalis* and *X. laevis* and each *Ngaro-like* family of *X. tropicalis* were used to compose the *DIRS-like* and *Ngaro-like* sequence libraries of each species. In the case of the *Ngaro-like* elements of *X. laevis*, as the families were not available in the Repbase, the consensus sequences were obtained from the copies recovered in the genomic search described above. For that, the sequences were aligned using MAFFT v7.471, with the pairwise divergence being assessed using Genedoc 2.7 (Nicholas et al., 1997) separating the sequences with more than 80% divergence into distinct groups (considering as different families). The consensus sequence of each group was obtained using UGENE (simple extended algorithm) with a 50% threshold (Okonechnikov et al. 2012). The *DIRS-like* and *Ngaro-like* libraries of each species were used to screen the genomes using RepeatMasker 4.1.0 (with the “-s”, “-nolow”, “-no_is”, “-a”, and “-lib” options). RepeatMasker utility Perl scripts were used to summarize the output (script buildSummary.pl) and to calculate Kimura 2-Parameter (K) divergence with adjusted CpG (script calcDivergenceFromAlign.pl). The scatter plot graphs representing the repeat landscape were created using the Python Matplotlib-v3.3.2 (Hunter, J.D., 2007) and edited in Inkscape software.

The age of the copies was estimated based on the time since the divergence of the ancestral sequence (since the consensus of each family used in the RepeatMasker is an approximation of its ancestor) using the formula: $T = K / r$ (Jiang et al. 2002), where a divergence (K) was obtained as described above, and r is the nucleotide substitution

rate of 3.1×10^{-9} substitutions per year, which is the average of the estimated substitution rates for *X. tropicalis* and *X. laevis*, and for the L and S subgenera of *X. laevis* (Session et al. 2016).

In order to investigate which families are being expressed in *X. tropicalis* and *X. laevis*, the expressed sequence tags (EST) library from both species were retrieved from Xenbase (<http://www.xenbase.org/>, RRID:SCR_003280) and the different DIRS families of both species were used as queries in BLASTn. The results were filtered by identity (>85%) and size (>100 bp).

RESULTS

Xenopus DIRS sequences belong to DIRS-like and Ngaro-like families

A total of 75 YR-retroelement families were identified in the Rebbase for *X. tropicalis* and 20 for *X. laevis* (Table 1). In this database, all YR-containing elements are classified as DIRS, a final group within the long terminal repeat (LTR) retrotransposons group. We thus previously assume that these families could belong to any of the DIRS superfamilies, then our analyses indicate they belong only to *DIRS-like* and *Ngaro-like*.

The evolutionary tree based on the RT domain of all the elements together recovered two well-supported clades (Figure 1) in which all the *Xenopus* sequences grouped in either (i) a *DIRS-like* or (ii) a *Ngaro-like* group. The two *PAT-like* sequences were not grouped as a monophyletic group (Figure 1).

We recognized only two of the 95 *Xenopus* Rebbase DIRS families as belonging to the *Ngaro-like* superfamily, i.e., DIRS-53_XT and DIRS-54_XT. Sequences from the families DIRS-6A_XT, DIRS-13C_XT, DIRS-27A_XT, DIRS-35_XT, DIRS-42_XT,

DIRS-3_XL, DIRS-10_XL, DIRS-13_XL, DIRS-15_XL, DIRS-18_XL, and DIRS-19_XL were not included in the tree because the RT domain was too short, but all these elements present a *DIRS-like* terminal repeat pattern.

The sequences of both superfamilies were analyzed, and a high level of congruence was found in the sequence structure in comparison with the DIRS families described in vertebrates (Goodwin and Poulter, 2004) (Figure 2). These findings will be discussed below.

DIRS-like families

The consensus sequences of the *DIRS-like* superfamily found in the *X. tropicalis* and *X. laevis* genomes range from 4146 base pairs (bp) in DIRS-33_XT to 6224 bp in DIRS-2A_XT. We recognized three ORFs in almost all the families with several levels of overlap, involving primarily ORF2 and ORF3 (see Table 1 and Supplementary Table 1, for more details).

The ORF1 encodes a gag-like protein and a LAP2alpha domain (approximately 650 amino acids) was predicted in all the families. The ORF2 corresponds to the RT and RH domains, with around 120 and 356 aa, respectively. A deoxy-adenosine methylase (DAM/MT) domain of around 284 aa was also observed in the ORF2 of some of the families evaluated here (DIRS-9A_XT, DIRS-9_XT, DIRS-10_XT, DIRS-16_XT, DIRS-22A_XT, DIRS-22_XT, DIRS-25_XT, DIRS-26A_XT, DIRS-26_XT, DIRS-27A_XT, DIRS-27_XT, DIRS-29_XT, DIRS-31_XT, DIRS-32_XT, DIRS-34A_XT, DIRS-34_XT, DIRS-35_XT, DIRS-37A_XT, DIRS-37_XT, DIRS-38_XT, DIRS-41_XT, DIRS-41A_XT, DIRS-42_XT, DIRS-46_XT, DIRS-50_XT, DIRS-51_XT, DIRS-52_XT, DIRS-1_XL, DIRS-3_XL, DIRS-4_XL, DIRS-5_XL, DIRS-

6_XL, DIRS-7_XL, DIRS-8_XL, DIRS-9_XL, DIRS-11_XL, DIRS-13_XL and DIRS-16_XL). The ORF3 encodes the YR protein with a conserved DNA_BRE_C domain of around 584 aa (Supplementary Table 1).

The *DIRS-like* elements have 5' and 3' ITRs and an ICR region (Figure 2A), and this pattern of repetition was found in almost all the *Xenopus DIRS-like* families evaluated here. The ITRs have approximately 120 bps and present a few nucleotide substitutions or *indels* between the left ITR (lITR) and the right ITR (rITR) (Figure 2A). The ICR is composed of two short sequences (lICR and rICR) which are complementary to the 5' (lTer) and the 3' (rTer) ends of the element (Figure 2A). The ICR and ITR sequences overlap slightly in most families.

Overall, 36 of the 75 Rebpase *DIRS-like* families of *X. tropicalis* present some level of degeneration in the molecular structure of the terminal repeats and/or ORFs domains (Table 1). In *X. laevis*, 13 of the 20 families present premature interruptions in the ORFs or incomplete repeats (Table 1), which indicates a high level of degeneration in these families (Supplementary Table 1). Concerning the EST data, we observed that most families present transcripts (55 families from *X. tropicalis* and 12 families from *X. laevis*) (Table 1). The *DIRS-like* families of both genomes have characteristic thymine trinucleotides (i.e., “TTT”) in both their 5' and 3' ends.

The RT sequence tree highlights the high level of diversity of the families of the *DIRS-like* clade in *Xenopus* (Figure 1). The diagnostic *DIRS-like* sequences (DIRS-1_ACar and DIRS-5_CBP) were recovered in an internal clade of the *Xenopus* sequences, which indicates that a number of *DIRS-like* families were already present in the ancestor of these species and thus contributed to the high current diversity of families.

As the Rebpase nomenclature of the families established for a species follows the order of their description (Kapitonov and Jurka, 2008; Bao et al. 2009), the evolutionary relationships among the families must be interpreted based on their phylogenetic relationships in the sequence trees, rather than their nomenclature in the databases. For example, DIRS-4_XT is not closely related to DIRS-4_XL, whereas DIRS-2_XL and DIRS-50_XT have a very close relationship. We recovered families of *DIRS-like* that were shared between the two species, such as DIRS-29_XT with DIRS-17_XL + DIRS-14_XL, DIRS-2_XL and DIRS-50_XT, DIRS-6_XL + DIRS-11_XL with DIRS-52_XT + DIRS-16_XT, and DIRS-41_XT + DIRS-41A_XT with DIRS-9_XL + DIRS-16_XL.

We also observed marked species-specific structuring in the sequence tree, recovering subclades that grouped families only from *X. tropicalis* or *X. laevis*. This indicates that many of the families may have originated after the separation of the two species, in particular in *X. tropicalis*. The landscape profile observed in each genome further reinforces this conclusion (Figure 3). The *X. laevis* landscape presents two broad ancient waves of invasion (of 60–90 million years ago - mya and 110–135 mya), whereas, in the *X. tropicalis* genome, there is a peak of very recent amplification, which occurred less than 3.2 mya. The most recent families of the *X. laevis* genome are found in smaller proportions, approximately 0.003% of the genome. Although the diversity of the *DIRS-like* families is much lower in *X. laevis*, they make up a larger proportion of the genome (0.8%) than in *X. tropicalis* (0.5%).

From potentially active families, we chose to analyze the DIRS-37_XT copies in the genomes. In *X. tropicalis*, we can observe that most copies recovered (except copies 4 and 6) are putative functional presenting all ORFs and conserved ITRs and ICR. Copies 1 to 7 are the most closely related and grouped in the clade containing DIRS-

37_XT, DIRS-37A_XT and DIRS-28_XT, but only the copies 1, 2 and 4 grouped with the query. The search has also recovered copies that grouped with more distant families, DIRS-31_XT (copies 8 and 9) and DIRS-3_XT (copy 10).

In *X. laevis*, the closest described families to DIRS-37_XT are DIRS-14_XL, DIRS-17_XL and DIRS-2_XL. All the copies recovered from the *X. laevis* genome present some level of degeneration, and due to this, copies 5 and 7 were not included in our phylogenetic inference. The nearest copy to the query is D37c1_XL, which was placed within a clade containing some *X. tropicalis* and no *X. laevis* families. This suggests that this sequence may be a degenerate copy of a *X. laevis* family that has yet to be established and was not successful in amplifying in this genome. Copies 2, 6, 8, and 9 grouped with DIRS-2_XL and DIRS-50_XT, while copies 3, 4, and 10 grouped with DIRS-14_XL, which is closely related to DIRS-17_XL and DIRS-29_XT. We found copies of the conserved elements in both subgenomes.

Ngaro-like families

The *Ngaro-like* families DIRS-53_XT and DIRS-54_XT also present the three expected ORFs (encoding *gag*-like elements, RT/RH, and YR) with an additional ORF encoding a protein containing the SGNH_hydrolase domain. The reading frames of all the ORFs overlap (Figure 2B; Supplementary Table 1). The *Ngaro-like* elements are known to have a different type of terminal repeat, the SDRs (Goodwin and Poulter, 2004). This pattern can be observed in the DIRS-53_XT copy (Figure 2B). This element has type A1 and A2 SDRs of 236 bps and B1 and B2 types of 152 bps, with both repetitions being 100% identical. No SDRs were found in the DIRS-54_XT. Although none of the *X. laevis* Rebbase DIRS families grouped with *Ngaro-like* clade, our

searches in the genome revealed the presence of sequences homologous to DIRS-53_XT and DIRS-54_XT, *Ngaro-like* transcripts were found for both species.

The copies of DIRS-53_XT and DIRS-54_XT recovered from *X. tropicalis* genome varied considerably in the degree of conservation of their sequences and structure (Supplementary Table 1), ranging from well-conserved copies to highly degenerate ones, due primarily to the loss of all or part of their 5' and/or 3' SDRs. Copies 1, 2, 4, 8, and 9 of DIRS-53_XT and copies 5 and 10 of DIRS-54_XT are potentially active.

In *X. laevis*, by contrast, we observed a high level of degeneration of the copies homologous to DIRS-53_XT/DIRS-54_XT, including a complete or partial loss of the 5' and/or 3' SDRs (Supplementary Table 1). These copies also present a high level of degeneration in the ORFs, including copies that lack the YR domain, hydrolase or RT domains (Supplementary Table 1). Of all copies recovered, only D54c5XL is potentially active.

The copies recovered from both species were included in the sequence tree (Figure 1), except those with a short RT domain (D53c2XT, D53c5XT, D53c9XT, D53c1XL, D53c2XL, D53c3XL, D53c4XL, D53c6XL, D53c7XL, D53c10XL, D54c1XL, D54c3XL, D54c6XL, D54c7XL, D54c9XL and D54c10XL). Clearly, DIRS-53_XT and DIRS-54_XT are very divergent (approximately 65% over only 600 bp of good alignment) and form two major groups of *Ngaro-like* sequences and some recovered copies did not group with the query. The DIRS-53 clade grouped copies retrieved with this query from *X. tropicalis* (copies 1, 3, 4, 6, 7, and 8) and copy 9 from *X. laevis*. The D54c5XL and D54c8XT copies also grouped with these sequences. This result suggests that this family is undergoing a process of degeneration in *X. laevis*, while it is still active in *X. tropicalis*.

The DIRS-54 clade contains the canonical element from the Rebbase, together with seven copies (1, 2, 3, 4, 5, 6 and 10) from the *X. tropicalis* genome (Figure 1). While the SDRs are not present in the consensus canonical copy, they can be observed in copies 5 and 10, which indicates that this family is also potentially active in *X. tropicalis*. None of *X. laevis* copies are grouped in this clade. The remaining copies from both species grouped in separate small clades and the copy D53c5XL is branching out from the base of the *Ngaro-like* clade, all of which are putative degenerate copies, given that none of them have SDRs. These sequences probably represent degenerate copies of the DIRS-53 or DIRS-54 families.

Contrary to what we observed for *DIRS-like*, *Ngaro-like* elements have similar densities in *X. tropicalis* and *X. laevis* of genomes, representing 0.049% and 0.047%, respectively. The *Ngaro-like* landscape profile shows a recent amplification signal (less than 3.2 mya) in the *X. tropicalis* and *X. laevis* genomes. Additionally, in both species it is possible to observe a very ancient amplification wave, dating from 120-145 mya, that is more prominent in *X. tropicalis* graph.

DISCUSSION

In this work, we presented a detailed analysis of the YR-containing elements (order DIRS) available in the Repbase23.11 for *X. tropicalis* and *X. laevis* classifying them as either into the *DIRS-like* or *Ngaro-like* superfamilies. We have provided a detailed description of the structural characteristics of these elements and found the specific molecular signature of each superfamily. *DIRS-like* and *Ngaro-like* elements present distinct diagnostic features in their 5' and 3' non-coding terminal regions, with the same structural elements already recorded in other metazoans, fungi, and protists

(Goodwin and Poulter 2001; Goodwin and Poulter 2004; Poulter and Goodwin 2005; Goodwin and Poulter 2015).

Our detailed analysis of the elements corroborates what has been described for the superfamilies also concerning the ORFs and domains (Goodwin and Poulter, 2001; Poulter and Goodwin, 2005). Thus, based on their molecular structure and phylogenetic criteria, only DIRS-53_XT_and DIRS-54_XT belong to the *Ngaro-like* superfamily, and all others are *DIRS-like*. We also identified conserved thymine trinucleotides (“TTT”) at the 5’ and 3’ ends in the complete copies of the *DIRS-like* elements of the genomes. This may be a general pattern of the *DIRS-like* elements found in terminal regions, which may be essential for transposing these elements (Malicki et al 2020). A similar terminal signature has also been described in the case of the DrDIRS1 element of *Tribolium castaneum*, which has either the trinucleotides “GTT” or dinucleotides “AA” (Goodwin et al. 2004). Given this, the recognition of the similarities in the molecular structure of these elements from the two species would also contribute to the assessment of the diversity and evolutionary history of these YR retrotransposons in other anuran genomes.

We found a much greater diversity and proportion of elements of the *DIRS-like* superfamily, in comparison with the *Ngaro-like* elements, in both genomes. *DIRS-like* and *Ngaro-like* elements are widely distributed in a number of metazoan groups, including fish, amphibians, reptiles, and in some fungus (Ruiz-Peres et al, 1996) and protists (Goodwin and Poulter, 2001; Poulter and Goodwin, 2005). Although there is no clear report of DIRS elements in some groups of species, we believe that the distribution of these retrotransposons is underestimated since they are frequently classified in the amount of the LTR retrotransposons group in the annotation of repetitive sequences based on the RepeatMasker tool.

The richness of the *DIRS-like* elements described in the present study had already been reported for the *X. tropicalis* genome (Hellsten et al. 2010; Piednoël et al. 2011) and reflects the importance of this repetitive element for its evolution. We recorded a much higher number of families of *DIRS-like* elements in *X. tropicalis* (81) than in the fish *Danio rerio* (12 families) or the reptile *Anolis carolinensis*, which has 42 families (Piednoël et al. 2011). We also found a large number of families are structurally complete in *X. tropicalis* and *X. laevis*. Thus, it is relevant to compare the landscape pattern of these superfamilies in both genomes since these species have undergone distinct evolutionary processes after the split. Session et al (2016) date the speciation of *X. tropicalis* and the *X. laevis* ancestor at around 48 mya. The *X. laevis* had an allotetraploid origin (around 17-18 mya) from two extant diploid progenitors separated at around 34 mya, and currently has two homoeologous subgenomes (L and S). The L and S subgenomes have undergone profound intragenomic diversification, which is compatible with the absence of recombination between the homeologous chromosome pairs of each subgenome since the allotetraploidization event (Session et al., 2016). The S subgenome has undergone intrachromosomal rearrangements and extensive small-scale deletions that resulted in the reduction of the length of the S chromosomes in comparison with their L homeologs (Sessions et al 2016).

If we consider the divergence time and evolutionary rates estimated by Session et al. (2016) as the most realistic scenario of the evolutionary history of *Xenopus* genus, our time-scale estimate of DIRS evolution in *X. tropicalis* and *X. laevis* genomes showed different patterns in both genomes. For *DIRS-like* it is possible to suggest that two ancient waves of amplification have occurred in the ancestor of both species. These ancient waves were less clear in *X. tropicalis* genome suggesting a higher turnover of degenerate copies in *X. tropicalis* than in *X. laevis* after the separation of these species.

Independent waves of *DIRS-like* amplification have also occurred after the separation of both species. It is possible to observe two more recent and modest amplification waves of *DIRS-like* in *X. laevis*, one estimated around 19 mya, probably in one or both ancestor genomes, and a more recent wave has occurred after the allotetraploidization, around 9.6 mya. This is consistent with the observation that some families are still potentially active in *X. laevis*. In *X. tropicalis* the amplification waves were more prominent occurring around 16 mya and less than 3.2 mya. The existence of potentially active copies of *DIRS-like* identified here is in agreement with the evidence of transcriptional activity detected in the transcriptome data on *X. tropicalis* (Poulter and Butler, 2015) and the searches performed by us in the EST libraries of both species. The recent burst events in both species could also have contributed to the diversification of families in each genome, primarily in *X. tropicalis*, which is consistent with the strong species-specific clustering seen in the sequence tree.

These two types of evidence (landscape pattern and phylogenetic) highlight the relative success of the *DIRS-like* elements in the *Xenopus* genomes in comparison with the *Ngaro-like* superfamily that presents very low diversity and quantity. *Ngaro-like* had a very ancient amplification wave followed by a long period of senescence. Despite *Ngaro-like* elements have failed to increase copy number and diversify in *Xenopus*, some potentially active copies are found in both genomes what is consistent with the very recent amplification wave (less than 3.2 Mya) seen in both species. The recent amplification indicates that *Ngaro-like* copies were maintained as active and somehow silenced for a long period of *Xenopus* evolution or were recently reactivated or reintroduced in these genomes. It is not clear, however, the reason why the *Ngaro-like* did not achieve the same success as *DIRS-like*, but it could be related to the known differences in the transposition mechanism (Poulter and Goodwin, 2005) of both

superfamilies or to a possible variation in silencing efficiency by the host, while it could also be explained by chance. Still, evidence of the accumulation of DIRS-1 element on centromeres of the *D. discoideum* genome (Dubin et al. 2010; Malicki et al., 2017) could indicate their role during the centromeric heterochromatin biogenesis in this genome and open new perspectives to future evaluation about their biological significance in their success in *Xenopus* genomes.

The presence of very ancient and senescent DIRS copies in both genomes is consistent with “TE cycle life” of the genome (Kidwell and Lisch, 2001), in which ancient mobile elements may lose their autonomy and no amplifications occur, with the nucleotide sequences losing their identity, with the senescent elements eventually being deleted or becoming completely divergent. When we discriminated the *DIRS-like* and *Ngaro-like* families exclusive to each *X. laevis* subgenome, we observed a much greater diversity of families and copies in the L subgenome in comparison with the S subgenome (Supplementary Table 1). The identification of families exclusive of each subgenome in *X. laevis* was reported to Miniature Inverted-repeat Transposable Elements (MITE) XI-TpL_Harb and XI-TpS_Harb, exclusive from subgenome L and S, respectively and also support the absence of recombination between the homeologous chromosome pairs of each subgenome since the allotetraploidization event (Session et al. 2016). The staggered pattern of superfamily diversity observed here (*X. tropicalis* > *X. laevis* L subgenome > *X. laevis* S subgenome) appears to reflect the same evolutionary pattern observed in other protein-coding genes, repetitive DNAs, and the rates of tandem duplications observed in each genome or subgenome (Session et al. 2016). However, as we found sequences in *X. laevis* from families non-reported in the Repbase, we cannot discard the possibility that the diversity of DIRS in this species is somewhat larger.

Our findings also point to differential evolutionary dynamics of the YR retrotransposons of the *DIRS-like* order in the diploid (*X. tropicalis*) and allopolyploid (*X. laevis*) species studied here. The intragenomic behavior of the transposable elements depends on the balance between repression and expression, due to the need to avoid a large number of copies becoming a disadvantage for the genome (Bourque et al., 2018). The conservation of the molecular structure of these elements is related directly to these genetic mechanisms, which determine either an increase or loss of TE diversity, depending on the repertoire of TEs, during the genomic evolution of each lineage.

Acknowledgements

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/PROAP – Finance Code 001) for the scholarships provided to CBG. We acknowledge Xenbase (www.xenbase.org) for staging, gene expression resources, and genomic reference material consulted throughout this work. We especially thank Luciana Bolsoni Lourenço, Maríndia Deprá and Michelle Orane Schemberger for valuable comments on a previous version of the manuscript.

Table 1: Basic features of the DIRS families of *Xenopus tropicalis* (XT) and *X. laevis* (XL) deposited in Rebbase. The consensus sequence of each family was evaluated for

the presence/absence of complete conserved ORFs containing the expected domains and the presence of complete structure of repeats (ITRs and ICR for *DIRS-like* and SDRs for *Ngaro-like*). The presence of ESTs is also shown for each family.

Name	Size	Superfamily	Preserved ORFs	Conserved structure of repeats	EST
DIRS-1_XT	5556 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-2_XT	5992 bp	<i>DIRS-like</i>	No, 1 and 3 not preserved	Yes	Yes
DIRS-2A_XT	6224 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-3A_XT	5809 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-3_XT	5870 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-4A_XT	5718 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-4B_XT	5380 bp	<i>DIRS-like</i>	Yes	No	No
DIRS-4_XT	5616 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-5_XT	5682 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-6A_XT	5710 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-6_XT	5814 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-7_XT	5503 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-8A_XT	5835 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-8_XT	5898 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-9A_XT	5447 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-9_XT	5386 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-10_XT	5129 bp	<i>DIRS-like</i>	No, 3 not preserved	Yes	No
DIRS-11_XT	5426 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-12_XT	5162 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-	5629 bp	<i>DIRS-like</i>	Yes	Yes	Yes

13A_XT					
DIRS-	5825 bp	<i>DIRS-like</i>	Yes	Yes	No
13B_XT					
DIRS-	5716 bp	<i>DIRS-like</i>	Yes	Yes	Yes
13C_XT					
DIRS-13_XT	5777 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-14_XT	5878 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-15_XT	5270 bp	<i>DIRS-like</i>	No, 3 not preserved	Yes	No
DIRS-16_XT	5346 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-	5302 bp	<i>DIRS-like</i>	Yes	Yes	No
17A_XT					
DIRS-17_XT	5368 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-18_XT	5685 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-19_XT	5643 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-20_XT	5075 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-	5716 bp	<i>DIRS-like</i>	Yes	Yes	Yes
21A_XT					
DIRS21_XT	5083 bp	<i>DIRS-like</i>	Yes	No	No
DIRS-	5059 bp/ 5298 bp*	<i>DIRS-like</i>	No, 2 not preserved	No	No
22A_XT	bp*		preserved		
DIRS-22_XT	5608 bp	<i>DIRS-like</i>	No 3, not preserved	Yes	Yes
DIRS-23_XT	5277 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-24_XT	5211 bp	<i>DIRS-like</i>	No, 3 not preserved	Yes	No
DIRS-	5316 bp	<i>DIRS-like</i>	Yes	No	No
25A_XT					
DIRS-25_XT	5504 bp	<i>DIRS-like</i>	Yes	Yes	Yes

DIRS-26A_XT	5370 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-26_XT	5415 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-27A_XT	5061 bp	<i>DIRS-like</i>	No, 3 not preserved	No	Yes
DIRS-27_XT	5611 bp	<i>DIRS-like</i>	No, 2 not preserved	No	Yes
DIRS-28_XT	5644 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-29_XT	5665 bp	<i>DIRS-like</i>	No, 1 not preserved	Yes	Yes
DIRS-30_XT	5781 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-31_XT	5691 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-32_XT	5187 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-33_XT	4146 bp	<i>DIRS-like</i>	No, 2 not preserved	Yes	Yes
DIRS-34A_XT	5490 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-34_XT	2880 bp/4728 bp*	<i>DIRS-like</i>	No, 2 not preserved	No	No
DIRS-35_XT	5900 bp	<i>DIRS-like</i>	No, 2 not preserved	No	Yes
DIRS-36A_XT	5595 bp	<i>DIRS-like</i>	No, 2 not preserved	No	No
DIRS-36B_XT	5562 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-36_XT	5415 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-37A_XT	5123 bp/5602 bp*	<i>DIRS-like</i>	No, 3 not preserved	Yes	Yes
DIRS-37_XT	5643 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-38_XT	5221 bp	<i>DIRS-like</i>	Yes	Yes	Yes

DIRS-39_XT	5484 bp	<i>DIRS-like</i>	No, 3 not preserved	Yes	Yes
DIRS-40_XT	5448 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-41A_XT	2952 bp/5431 bp*	<i>DIRS-like</i>	No, 3 not preserved	No	No
DIRS-41_XT	5212 bp	<i>DIRS-like</i>	No, 3 not preserved	Yes	No
DIRS-42_XT	5710 bp	<i>DIRS-like</i>	No, 2 not preserved	No	Yes
DIRS-43_XT	5156 bp	<i>DIRS-like</i>	No, 2 not preserved	Yes	Yes
DIRS-44_XT	4762 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-45_XT	5054 bp	<i>DIRS-like</i>	Yes	No	Yes
DIRS-46_XT	5239 bp	<i>DIRS-like</i>	No, 3 not preserved	Yes	No
DIRS-47_XT	3912 bp/4490 bp*	<i>DIRS-like</i>	No, 3 not preserved	No	No
DIRS-48_XT	5153 bp	<i>DIRS-like</i>	No, 2 not preserved	Yes	No
DIRS-49_XT	5152 bp	<i>DIRS-like</i>	No, 2 not preserved	No	Yes
DIRS-50_XT	5778 bp	<i>DIRS-like</i>	No, 3 not preserved	No	Yes
DIRS-51_XT	4911 bp	<i>DIRS-like</i>	No, 1 and 2 not preserved	No	Yes
DIRS-52_XT	5533 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-53_XT	5457 bp	<i>Ngaro-like</i>	Yes	Yes	Yes
DIRS-54_XT	5141 bp	<i>Ngaro-like</i>	Yes	No	Yes
DIRS-1_XL	5291 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-2_XL	5742 bp	<i>DIRS-like</i>	No, 1 and 3	Yes	Yes

not preserved					
DIRS-3_XL	5387 bp	<i>DIRS-like</i>	No, 1 and 2	Yes	No
not preserved					
DIRS-4B_XL	5417 bp	<i>DIRS-like</i>	No, 2 not	Yes	Yes
preserved					
DIRS-4_XL	5433 bp	<i>DIRS-like</i>	No, 3 not	Yes	Yes
preserved					
DIRS-5_XL	5464 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-6_XL	5456 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-7_XL	5470 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-8_XL	5440 bp	<i>DIRS-like</i>	Yes	Yes	No
DIRS-9_XL	5226 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-10_XL	5852 bp	<i>DIRS-like</i>	No, 3 not	No	Yes
preserved					
DIRS-11_XL	5397 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-12_XL	5567 bp	<i>DIRS-like</i>	No, 1 not	Yes	Yes
preserved					
DIRS-13_XL	5531 bp	<i>DIRS-like</i>	No, 3 not	Yes	Yes
preserved					
DIRS-14_XL	5569 bp	<i>DIRS-like</i>	Yes	Yes	Yes
DIRS-15_XL	5660 bp	<i>DIRS-like</i>	No, 1 and 3	No	Yes
not preserved					
DIRS-16_XL	5565 bp	<i>DIRS-like</i>	No, 3 not	Yes	No
preserved					
DIRS-17_XL	5711 bp	<i>DIRS-like</i>	No, 1 not	Yes	Yes
preserved					
DIRS-18_XL	4918 bp	<i>DIRS-like</i>	No, 1,2 and 3	No	Yes
not preserved					
DIRS-19_XL	5064bp	<i>DIRS-like</i>	No, 3 not	Yes	No
preserved					

Note: * - Rebbase size/recovered copies size – for some families, the structure of terminal repeats was absent in the Rebbase consensus sequences and was further analyzed by recovering copies from the genomes.

Figure legends:

Fig. 1 Sequence tree produced by Bayesian inference, based on the amino acid sequences of the reverse transcriptase domain. The matrix was composed of the sequences of *X. tropicalis* and *X. laevis* DIRS- elements obtained from the Rebbase database, the copies retrieved from both genomes and diagnostic sequences from each DIRS superfamily. The posterior probabilities are indicated at the branches. Sequences of different superfamilies are highlighted in different colors: blue for *Ngaro-like*, green for *PAT-like*, and red for *DIRS-like*. The shades of each color also distinguish the sequences of *X. tropicalis* (dark) and *X. laevis* (light)

Fig. 2 Schematic structure of the potentially complete *DIRS-like* and *Ngaro-like* retroelements of the *Xenopus tropicalis* genome. (A) Representation of the *X. tropicalis* *DIRS-like* elements, based on the Rebbase consensus sequences of the DIRS-37_XT, that contain three ORFs, conserved domains (gag, RT/RH/MT, and YR) and non-coding portions: inverted terminal repeats (ITRs) and internal complementary regions (ICRs). The expanded scheme of the terminal regions is shown in the lower plot. (B) Representation of the *X. tropicalis* *Ngaro-like* elements based on the Rebbase consensus sequences of the DIRS-53_XT that contain four ORFs, conserved domains (RT/RH, YR, and SGNH) and split direct repeats (SDRs)

Fig. 3 Evolution of *Xenopus* species and evolutionary dynamics of DIRS elements. A) Evolutionary events as estimated by Session et al (2016). The speciation of *X. tropicalis* and the ancestor of *X. laevis* at 48 mya, the speciation of the L and S progenitors at 34 mya, and the hybridization around 17-18 mya. B) The graphs show the divergence of *DIRS-like* (above) and *Ngaro-like* (below) copies mapped in the genomes of *X. tropicalis* (red) and *X. laevis* (green) with their consensus sequence expressed in Kimura-2-parameters distance and the corresponding time of divergence in million years (x-axis) plotted in relation to the proportion in the genome (y-axis).

REFERENCES

- Bao W, Jurka M G, Kapitonov V V, Jurka J (2009). New Superfamilies of Eukaryotic DNA Transposons and Their Internal Divisions. *Molecular Biology and Evolution*, 26(5), 983–993. doi:10.1093/molbev/msp013
- Bourque G, Burns K H, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin H L, Macfarlan T S, Mager D L, Feschotte C (2018). Ten things you should know about transposable elements. *Genome Biology*, 19(1), 199–. doi:10.1186/s13059-018-1577-z
- Bowes J B, Snyder K A, Segerdell E, Jarabek C J, Azam K, Zorn A M, Vize P D (2010) Xenbase: Gene expression and improved integration. *Nucleic Acids Research*, v. 38, D607–612
- Cappello J, Handelsman K, Lodish H F (1985) Sequence of Dictyostelium DIRS-1: An apparent retrotransposon with inverted terminal repeats and an internal circle junction

sequence. *Cell*, v. 43, n. 1, p. 105–115.

Dasilva C, Hadji H, Ozouf-Costaz C, Nicaud S, Jaillo, O, Weissenbach J, Crollius H R. (2002). Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome. *Proceedings of the National Academy of Sciences*, 99(21), 13636–13641. doi:10.1073/pnas.202284199

Dubin M., Fuchs J., Gräf R., Schubert I., Nellen W (2010). Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Research*, 38(21), 7526–7537. doi:10.1093/nar/gkq664

Duncan L, Bouckaert K, Yeh F, Kirk D (2002) kangaroo, a mobile element from *Volvox carteri*, is a member of a newly recognized third class of retrotransposons. *Genetics*. Dec;162(4):1617-30.

Edwards R J, Tuipulotu D E, Amos T G, O’Meally D, 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*. Doi:10.1093/gigascience/giy095

Goodwin T J D, Poulter R T M (2001) The DIRS1 group of retrotransposons. *Molecular Biology and Evolution*, v. 18, n. 11, p. 2067–2082.

Goodwin T J D, Poulter R T M (2004) A New Group of Tyrosine Recombinase-Encoding Retrotransposons. *Molecular Biology and Evolution*, v. 21, n. 4, p. 746–759.

Goodwin T J D, Poulter R T M, Lorenzen M D, Beeman R W (2004) DIRS retroelements in arthropods: Identification of the recently active TcDirsl element in the red flour beetle *Tribolium castaneum*. *Molecular Genetics and Genomics*, v. 272, n. 1, p. 47–56.

Hammond S A, Warren R L, Vandervalk B P, Kucuk E, Khan H, Gibb E A, ... Birol I

(2017). The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nature Communications*, 8(1). doi:10.1038/s41467-017-01316-7

Hellsten U, Harland R M, Gilchrist M J, Hendrix D, Jurka J, Kapitonov V, ... Rokhsar D S (2010). The Genome of the Western Clawed Frog *Xenopus tropicalis*. *Science*, 328(5978), 633–636. doi:10.1126/science.1183670

Hunter J D (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95 doi:10.1109/mcse.2007.55.

Jiang, N. (2002). Dasheng and RIRE2. A Nonautonomous Long Terminal Repeat Element and Its Putative Autonomous Partner in the Rice Genome. *PLANT PHYSIOLOGY*, 130(4), 1697–1705. doi:10.1104/pp.015412

Jiang L, Ruan Q, Chen W. (2015). The complete mitochondrial genome sequence of the Xizang Plateau frog, *Nanorana parkeri*(Anura: Dicroglossidae). *Mitochondrial DNA*, 1–2. doi:10.3109/19401736.2015.1007327

Jurka J (2000) Rebbase Update: A database and an electronic journal of repetitive elements. *Trends in Genetics*, v. 16, n. 9, p. 418–420.

Kapitonov V V, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Rebbase. *Nat Rev Genet*, v. 9, n. 5, p. 411–2.

Kidwell M G, Lisch D R (2001). Perspective: Transposable Elements, Parasitic Dna, And Genome Evolution. *Evolution*, 55(1), 1–24. Doi:10.1111/j.0014-3820.2001.tb01268.x

Kohany O, Gentles A J, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Rebbase: RebbaseSubmitter and Censor. *BMC Bioinformatics*, v. 7, p. 1–7.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular

evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, v. 35, n. 6, p. 1547–1549.

Kuznetsova I S, Thevasagayam N M, Sridatta P S R, Komissarov A S, Saju J M, Ngoh S Y, Shen X, Jang J, Órban L (2014). Primary analysis of repeat elements of the Asian seabass (*Lates calcarifer*) transcriptome and genome. *Frontiers in Genetics*, 5. doi:10.3389/fgene.2014.00223

Letunic, I.; Bork, P (2019) Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, v. 47, n. W1, p. 256–259. Oxford University Press.

Malicki, M; Iliopoulou, M; Hammann, C (2017). Retrotransposon Domestication and Control in *Dictyostelium discoideum*. *Frontiers in Microbiology*, 8(), 1869–. doi:10.3389/fmicb.2017.01869

Malicki M, Spaller T, Winckler T, Hammann C (2020). DIRS retrotransposons amplify via linear, single-stranded cDNA intermediates. *Nucleic Acids Research*. doi:10.1093/nar/gkaa160

Marchler-Bauer A, Bryant S H (2004) CD-Search: Protein domain annotations on the fly. *Nucleic Acids Research*, v. 32, n. WEB SERVER ISS., p. 327–331.

Murphy W J (2005). Dynamics of Mammalian Chromosome Evolution Inferred from Multispecies Comparative Maps. *Science*, 309(5734), 613–617. doi:10.1126/science.1111387

Nicholas K B, Nicholas Jr H B (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Available at: <http://www.nrbsc.org/gfx/genedoc/gdsrc.htm>.

Notredame C, Higgins D G, Heringa, J (2000) T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, v. 302, n. 1, p.

205–217.

Okonechnikov K, Golosova O, Fursov M. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28(8), 1166–1167. doi:10.1093/bioinformatics/bts091

Piednoël M, Gonçalves I R, Higuier D, Bonnivard, E (2011) Eukaryote DIRS1-like retrotransposons : an overview. , p. 1–18.

Poulter R T M, Goodwin T J D (2005). DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenetic and Genome Research*, 110(1-4), 575–588. doi:10.1159/000084991

Poulter R T M, Butler M I (2015) Tyrosine Recombinase Retrotransposons and Transposons. *Mobile DNA III*, p. 1271–1291.

Poulter R T M, Goodwin T J D (2005) DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenetic and Genome Research*, v. 110, n. 1–4, p. 575–588.

Ribeiro Y C, Robe L J, Veluza D S, dos Santos C M B, Lopes A L K, Krieger M A, Ludwig A (2019). Study of VIPER and TATE in kinetoplastids and the evolution of tyrosine recombinase retrotransposons. *Mobile DNA*, 10(1). doi:10.1186/s13100-019-0175-2

Ronquist F, Teslenko M, van der Mark P, Ayres D L, Darling A, Höhna S, ... Huelsenbeck J P (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), 539–542. doi:10.1093/sysbio/sys029

Ruiz-Perez VL, Murillo FJ, Torres-Martinez S. (1996) Prt1 an unusual retrotransposon-like sequence in the fungus *Phycomyces blakesleeianus*. *Mol Gen Genet* 253:324–333.

Seidl F, Levis N A, Schell R, Pfennig D W, Pfennig K S, Ehrenreich I M (2019). Genome of *Spea multiplicata*, a Rapidly Developing, Phenotypically Plastic,

and Desert-Adapted Spadefoot Toad. Genetics, g3.400705.2019. doi:10.1534/g3.119.400705

Session A M, Uno Y, Kwon T, Chapman J A, Toyoda A, Takahashi S, ... Kondo M (2016). Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*, 538(7625), 336–343. doi:10.1038/nature19840

Wicker T, Sabot F, Hua-Van A, Bennetzen J L, Capy P, Chalhoub B, ... Schulman A H (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973–982. doi:10.1038/nrg2165

Wicker, Thomas; Gundlach, Heidrun; Spannagl, Manuel; Uauy, Cristobal; Borrill, Philippa; Ramírez-González, Ricardo H.; De Oliveira, Romain; Mayer, Klaus F. X.; Paux, Etienne; Choulet, Frédéric (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology*, 19(1), 103–. doi:10.1186/s13059-018-1479-0

A

B

LEGEND:

ITer- ITR	A1/A2	ORF tyrosine recombinase (YR)	Methyltransferase domain (MT)
rTer - ITR	B1/B2	ORF reverse transcriptase (RT)/ RNase H (RH)
IICR/rICR	ORFgag-like (GAG)	ORF SGNH_hydrolase (SGNH)	

A

B

DIRS – like *Ngaro – like*