UNIVERSIDADE FEDERAL DO PARANÁ

CAMILLA BORGES GAZOLLA

ESTRUTURA MOLECULAR, DIVERSIDADE E EVOLUÇÃO DOS RETROTRANSPOSONS DA ORDEM DIRS NO GENOMA DE TRÊS ESPÉCIES DA FAMÍLIA PIPIDAE (ANURA)

> CURITIBA 2022

CAMILLA BORGES GAZOLLA

ESTRUTURA MOLECULAR, DIVERSIDADE E EVOLUÇÃO DOS RETROTRANSPOSONS DA ORDEM DIRS NO GENOMA DE TRÊS ESPÉCIES DA FAMÍLIA PIPIDAE (ANURA)

Tese apresentada ao curso de Pós-graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná (UFPR), como requisito parcial à obtenção do título de Doutora em Genética.

Orientador: Prof Dr Daniel Pacheco Bruschi Coorientadora: Profa Dra Adriana Ludwig

CURITIBA 2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP) UNIVERSIDADE FEDERAL DO PARANÁ SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS BIOLÓGICAS

Gazolla, Camilla Borges

Estrutura molecular, diversidade e evolução de retrotransposons da ordem DIRS no genoma de três espéceis da Família Pipidae (Anura) / Camilla Borges Gazolla. – Curitiba, 2022.

1 recurso on-line : PDF.

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Biológicas, Programa de Pós-Graduação em Genética. Orientador: Prof Dr Daniel Pacheco Bruschi. Coorientadora: Profa Dra Adriana Ludwig.

1. Anuro. 2. Retroelementos. 3. Elementos de DNA Transponíveis. I. Bruschi, Daniel Pacheco, 1987-. II. Ludwig, Adriana. III. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética. IV. Título.

Bibliotecária: Giana Mara Seniski Silva. CRB-9/1406



MINISTÉRIO DA EDUCAÇÃO SETOR DE CIÊNCIAS BIOLÓGICAS UNIVERSIDADE FEDERAL DO PARANÁ PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -40001016006P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de CAMILLA BORGES GAZOLLA intitulada: ESTRUTURA MOLECULAR, DIVERSIDADE E EVOLUÇÃO DOS RETROTRANSPOSONS DA ORDEM DIRS NO GENOMA DE TRÊS ESPÉCIES DA FAMÍLIA PIPIDAE (ANURA), que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutora está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação. CURITIBA, 30 de Março de 2022.

> Assinatura Eletrônica 31/03/2022 08:34:03.0 DANIEL PACHECO BRUSCHI Presidente da Banca Examinadora

Assinatura Eletrônica 30/03/2022 22:36:19.0 LUCIANA BOLSONI LOURENÇO Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS) Assinatura Eletrônica 31/03/2022 08:24:57.0 MARCELO RICARDO VICARI Avaliador Interno (UNIVERSIDADE ESTADUAL DE PONTA GROSSA)

Assinatura Eletrônica 17/05/2022 06:26:07.0 PATRÍCIA PASQUALI PARISE MALTEMPI Avaliador Externo (UNIVERSIDADE ESTADUAL PAULISTA - CAMPUS DE RIO CLARO) Assinatura Eletrônica 30/03/2022 21:28:59.0 ADRIANA LUDWIG Coorientador(a) (INSTITUTO CARLOS CHAGAS)

Setor de Ciências Biológicas, Centro Politécnico - CURITIBA - Paraná - Brasil CEP 81531-980 - Tel: (41) 3361-1587 - E-mail: ppg-gen@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015. Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 170129

Para autenticar este documento/assinatura, acesse https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp e insira o codigo 170129

Dedico este trabalho aos meus pais, Liamar e Humberto, e a minha irmã Fabíolla.

AGRADECIMENTOS

Foi algo único a experiência de realizar o doutorado ao longo desses quatro anos. Foram tantos obstáculos, medos e incertezas, que chegar até aqui hoje, me dá uma sensação de alívio e gratidão. Portas precisam ser fechadas, para que outras sejam abertas. Comigo está sendo assim, quero fechar este ciclo de minha vida com a certeza que fiz o melhor que eu pude nas condições que eu tive. De uma forma ou de outra, poder contribuir, nem que seja um pouquinho, para a Ciência, é algo que todos deveriam fazer antes de partir. Eu entendo que nos dias atuais seja difícil manter a motivação perante as dificuldades, porém me sinto privilegiada por ter chegado até aqui. E digo que a base disso tudo, foi, é e sempre será, Deus.

Desejo exprimir minha gratidão a todos aqueles que estiveram comigo nesse tempo. As pessoas que vieram comigo, as pessoas que se foram e as pessoas que surgiram e ficaram. Em primeiro lugar quero agradecer ao Professor Daniel por ter confiado a mim este trabalho, sempre muito dedicado, solícito e disponível. Uma pessoa que inspira e motiva como pessoa e profissional, indo além das expectativas em tudo aquilo que se dedica a fazer. Também quero agradecer a Professora Adriana, que sempre com muita paciência e compreensão procura transmitir aquilo que sabe. É admirável a humildade, forma amiga e generosa que a Profa Adriana teve em nos ensinar e acompanhar todo o processo da bioinformática, pois no começo, eu não sabia nem o que era um *contig*.

Quero expressar todo o meu amor pela minha família, muito mais que a minha base, eles são meu porto seguro, mamãe Liamar, papai Humberto e irmã Fabíolla. Vocês são tudo para mim. Todas as minhas conquistas só fazem sentido se vocês estiverem comigo. Aproveito para salientar a importância de minhas avós (Maria e Nilza) e meus avôs (Adalto e Irineu – *in memorian*), que ajudaram a construir a mulher que sou hoje. Além disso, agradeço pelas alegrias que meus primos me proporcionam e toda energia positiva que sempre compartilham comigo.

Agradeço aos meus amigos de trabalho, que nunca mediram esforços para fazer as trocas de plantão sempre que eu precisei me dedicar ao doutorado. Sem vocês, eu não sei se teria conseguido.

Sou abençoada pelas amizades que tenho, pode ser que eu esqueça (neste momento) de agradecer alguém, mas meu coração tem um carinho especial por cada um que passou pela minha vida. Joana minha eterna dupla, como foi bom trabalhar com você, sinto saudades de rir contigo. Meus amigos Michelle, Jonny, Iraine, Júlia, Deborah, Amanda, Fernanda, Gyslaine, Ísis e Íris, foi muito bom esse tempo de laboratório e estarei sempre na torcida pelo sucesso de vocês.

Michelle e Juliana, vocês são minhas irmãs de alma, sou grata por cada dia estarmos mais conectadas e fortalecidas. Além disso, meus amigos, Marcela Tavares, Juliana Barros, Ana Libera, Arilúcio, Aline Domingues, Pâmela Regina, Bruna Colombo, Jéssica Monteiro, Ruan Bentes, Ramon, Edina, Sheila, Jana, Bene Felipe Burille e Eduardo Zózimo, por serem a parte leve de minha vida durante esse tempo, por me motivarem e me tirarem sorrisos. Vocês me proporcionaram boas conversas, aprendizados, cafés e companhia de forma a me fazer crescer como pessoa e creio que isso é essencial como parte do desenvolvimento de cada um.

Trik, é até difícil falar de você. Sou tão grata. Você é uma das pessoas que mais admiro. Um profissional exemplar e um ser humano ainda melhor. Que bom eu tê-lo em minha vida. Que bom ter tido a oportunidade de trabalhar com você e hoje podê-lo chamá-lo de amigo.

Sou grata pelos profissionais que me acompanharam durante esse tempo, Valdete e Dr Luiz Felipe, vocês foram essenciais na construção e descobrimento da pessoa que sou hoje. Além de cuidar da mente, agradeço ao meu amigo Jefferson Eich que me acompanhou em muitos treinos, mostrando que o cuidado físico é de extrema importância no dia a dia tão corrido que vivemos. Vocês três foram o combo para que eu conseguisse finalizar o Doutorado com o corpo e a mente em equilíbrio.

Finalizo agradecendo à CAPES, CNPQ e Fundação Araucária pelo apoio financeiro destinado a pesquisa. Além disso, agradeço a todos os colaboradores do setor de ciências biológicas, desde as pessoas da manutenção até os professores, pela oportunidade de ter trabalhado no laboratório.

"Am I wrong for thinking out the box from where I stay? Am I wrong for saying that I'll choose another way?"

RESUMO

Elementos transponíveis (TEs - Transposable Elements) são elementos genéticos com capacidade de se transpor nos genomas em que estão inseridos. O mecanismo de transposição destes elementos separa os TEs em duas grandes classes, sendo os elementos de Classe I, genericamente conhecidos como retrotransposons e de classe II, transposons. Os TEs influenciam no funcionamento do genoma, evidências apontam que esses elementos têm contribuido diretamente para arquitetura, manutenção e evolução desses genomas, os tornando dinâmicos. Sendo assim, a caracterização dos TEs, é um importante aspecto no estudo da evolução do genoma e por consequência, evolução dos cariótipos. A caracterização dos TEs de um genoma, incluem análise da estrutura e conservação de cópias, classificação e análise da sua história evolutiva, permitindo em muitos casos identificar o tipo de transferência, seja ela vertical ou horizontal, durante a evolução do genoma. O mapeamento dos TEs, aparece como uma ferramenta da biologia molecular adicional, favorecendo evidências sobre sua participação durante a evolução cromossômica. A classe Amphibia tem como principal característica os animais que possuem o ciclo de vida dividido em uma fase aquática e uma terrestre. Dentre as famílias da ordem Anura destaca-se a família Pipidae, que abrange 41 espécies divididas entre 4 gêneros. As espécies do gênero Pipa são anfíbios exclusivamente aquáticos e se localizam principalmente em locais de águas permanentes. Devido a proximidade filogenética de Pipa e Xenopus, e do fato de que Xenopus tropicalis foi o primeiro genoma sequenciado de anfíbios, isso possibilita a sua utilização como genoma de referência em estudos de genômica comparativa e evolução cromossômica em relação ao genoma de Pipa carvalhoi, recém sequenciado por nosso grupo de pesquisa. Com a análise do cariótipo de Xenopus em relação ao de Pipa carvalhoi, podemos propor algumas homologias cariotípicas primárias, baseadas em seu número diplóide e na morfologia dos cromossomos. Combinar ferramentas genômicas e de citogenética são uma forma arrojada de investigar a participação desses elementos na diversificação dos cariótipos e podem trazer importantes contribuições no estudo da biologia cromossômica.

Palavras-chaves: Elementos transponíveis, DIRS, NGARO, Anuros

ABSTRACT

Transposable Elements (TEs) are genetic elements capable of being transposed into the genomes in which they are inserted. The transposition mechanism of these elements separates TEs into two major classes, Class I elements, generically known as retrotransposons, and Class II elements, transposons. TEs influence the functioning of the genome, evidence indicates that these elements have contributed directly to the architecture, maintenance and evolution of these genomes, making them dynamic. Thus, the characterization of TEs is an important aspect in the study of genome evolution and, consequently, karyotype evolution. The characterization of the TEs of a genome includes analysis of the structure and conservation of copies, classification and analysis of its evolutionary history, allowing in many cases to identify the type of transfer, whether vertical or horizontal, during the evolution of the genome. The mapping of TEs appears as an additional molecular biology tool, providing evidence on their participation during chromosomal evolution. The Amphibia class has as its main characteristic the animals that have a life cycle divided into an aquatic and a terrestrial phase. Among the families of the Anura order, the Pipidae family stands out, which comprises 41 species divided into 4 genera. Species of the genus *Pipa* are exclusively aquatic amphibians and are mainly located in permanent water sites. Due to the phylogenetic proximity of *Pipa* and *Xenopus*, and the fact that Xenopus tropicalis was the first sequenced amphibian genome, this allows its use as a reference genome in comparative genomic studies and chromosomal evolution in relation to the genome of *Pipa carvalhoi*, recently sequenced by our research group. With the analysis of the Xenopus karyotype in relation to that of Pipa carvalhoi, we can propose some primary karyotypic homologies, based on their diploid number and chromosome morphology. Combining genomic and cytogenetic tools is a bold way to investigate the participation of these elements in karyotype diversification and can bring important contributions to the study of chromosomal biology.

Keywords: Transposable elements, DIRS, NGARO, Anurans

LISTA DE FIGURAS

CAPÍTULO 1

Figure 1: Sequence tree produced by neighbor-joining method (JTT ensure G), 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 Rebpase database, the copies retrieved from both genomes and diagnostic sequences from each DIRS superfamily. The bootstrap values higher than 50 are indicated at the branches. The * sign near the nodes indicates the clade was supported with posterior probability higher than 80 in the Bayesian tree. Sequences of different superfamilies are highlighted in different colors and shades of each color also distinguish the sequences of X. tropicalis (XT) and X. laevis (XL).

Figure 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 Rebpase consensus sequences of the DIRS-37_XT, that contain three ORFs, conserved domains (gag, RT/ RH/MT, and YR) and noncoding 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 Rebpase consensus sequences of the DIRS-53_XT that contain four ORFs, conserved domains (RT/RH, YR, and SGNH) and split direct repeats (SDRs).

Figure 3: Evolution of Xenopus species and evolutionary dynamics of DIRS elements. The evolutionary events as estimated by Session et al. (2016) are shown: the speciation of X. tropicalis and the ancestor of X. laevis at 48 mya, the speciation of the L and S progenitors of X. laevis at 34 mya, and their hybridization around 17–18 mya. The graphs show the divergence of DIRS-like (above) and Ngaro-like (below) copies mapped in the genomes of X. tropicalis and X. laevis (S and L subgenomes) 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).

CAPÍTULO 2

Figura 1: Cariograma correspondente a X. tropicalis submetido ao método de FISH com sonda DIRS-37_XT (DIRS-like) marcada com digoxigenina-11-dUTP, na suspensão do tipo Speedy, contendo a trissomia do cromossomo 10. O padrão de marcação sempre foi observado em pelo menos dez metáfases para cada sonda.

Figura 2: Cariograma correspondente a X. tropicalis submetido ao método de FISH com sonda NGARO-53_XT (Ngaro-like) marcada com digoxigenina-11-dUTP, na suspensão do tipo Speedy, contendo a trissomia do cromossomo 10. O padrão de marcação sempre foi observado em pelo menos dez metáfases para cada sonda.

Figura 3: Metáfase de Pipa carvalhoi submetida à FISH com sonda DIRS-37_XT (DIRSlike) marcada com digoxigenina-11-dUTP. (A) cromossomos corados com DAPI, (B) detecção da sonda DIRS-37_XT marcada com digoxigenina-11-dUTP e (C) sobreposição das capturas de imagens. Note que não há sinal de marcações da sonda nos cromossomos.

Figura 4: Metáfase de Pipa carvalhoi submetida à FISH com sonda NGARO-53_XT (NGARO-like) marcada com digoxigenina-11-dUTP. (A) cromossomos corados com DAPI, (B) detecção da sonda NGARO-53_XT marcada com digoxigenina-11-dUTP e (C) sobreposição das capturas de imagens. Note que não há sinal de marcações da sonda nos cromossomos.

Figura 5: Metáfase correspondente a *X. tropicalis submetida à FISH* com sonda DIRS-53_XT (semelhante a *Ngaro-like*) marcada com digoxigenina-11-dUTP. As hibridizações correspondem às marcações na coloração rosa presente nos cromossomos.

Figura 6: Cromossomos metafásicos de P. carvalhoi, submetidos a experimentos de CGH. O genoma de P. carvalhoi foi marcado com digoxigenina-11-dUTP (vermelho), enquanto o genoma de X. tropicalis foi marcado com biotina (verde). As regiões sobrepostas são exibidas em amarelo nos cromossomos metafásicos de P. carvalhoi.

LISTA DE TABELAS

CAPÍTULO 1

Table 1: Basic features of the DIRS families of Xenopus tropicalis (XT) and X. laevis (XL) deposited in Rebpase. 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.

CAPÍTULO 2

Tabela 1: Conteúdo dos elementos de transposição no genoma de Pipa carvalhoi

Tabela 1 – material suplementar: Análise da estrutura molecular dos elementos de transposição da ordem DIRS no genoma de Pipa carvalhoi. A sequência consenso de cada família recuperada foi avaliada quanto à presença/ausência das ORFs contendo ou não ou domínios conservados e também a presença/ausência das repetições terminais.

SUMÁRIO

1. INTRODUÇÃO12
1.1 Elementos genéticos móveis: uma visão geral
1.2 Presença e atuação dos elementos genéticos móveis nos genomas
1.3 Elementos genéticos móveis nos genomas de Anuros e a ordem DIRS
1.4 Xenopus tropicalis, Xenopus laevis e Pipa carvalhoi, representantes da família
Pipidae, que é um interessante grupo para estudo da evolução cromossômica e
genômica
2. OBJETIVOS
2.1 Objetivos gerais
2.2 Objetivos específicos
3. CAPÍTULO 1 - Evolutionary dynamics of DIRS-like and Ngaro-like retrotransposons
in X <i>enopus laevis</i> and <i>Xenopus tropicalis</i> genomes 24
3.1 Abstract
3.2 Introduction
3.3 Material and Methods
3.4 Results
3.5 Discussion
3.6 Literature cited
4. CAPÍTULO 2 – Estudo comparativo das famílias DIRS-like e Ngaro-like no genoma
de Pipa carvalhoi e Xenopus Tropicalis com evidências de regiões repetitivas espécie-
específica em <i>X. tropicalis</i> 54
4.1 Resumo
4.2 Introdução
4.3 Materiais e Métodos
4.4 Resultados
4.5 Discussão
4.6 Referências
5. CONSIDERAÇÕES FINAIS75

INTRODUÇÃO

Elementos genéticos móveis: uma visão geral

Os elementos genéticos móveis ou elementos transponíveis (*Transposable elements* - TEs) foram descobertos por Barbara McClintok durante seus estudos em *Zea mays.* Esses elementos possuem como particularidade a capacidade de transposição nos genomas em que estão inseridos (McClintock, 1950). Com o desenvolvimento das tecnologias de sequenciamento em larga escala, ocorreu um aumento no número de genomas disponíveis para as análises genômicas, e isso têm demonstrado que os TEs são sequências repetitivas onipresentes e abundantes nos genomas de eucariotos (López-Flores e Garrido Ramos, 2012; Berthelier et al., 2018; Wells e Feschotte, 2020), atuando na estrutura, manutenção e evolução dos genomas desses organismos (Wicker et al., 2007; Prithan, 2009; Sotero-Caio et al., 2017).

Os TEs estão presentes nos mais diversos grupos de eucariotos (Pappalardo et al., 2021), por exemplo, mamíferos (Senft et al., 2021), plantas (Thieme e Roulin, 2021), fungos (Fouché et al., 2021), insetos (Gilbert et al., 2021), aves (Wang et al., 2021) e répteis (Ahmad et al., 2021). Os grupos que apresentam em seu conteúdo genômico os TEs, mostram uma variação interespecífica com relação as classes de TEs que estão presentes, bem como há uma variação em nível quantitativo no genoma (Papallardo et al., 2021).

Com o avanço dos estudos genômicos relacionados a esses elementos nos mais diversos organismos e a grande variedade estrutural encontrada entre esses elementos, o estabelecimento de critérios para classificá-los se fez necessário. Alguns sistemas de classificações foram desenvolvidos para melhorar a compreensão desses elementos, considerando principalmente o mecanismo de transposição do elemento, além das características estruturais e moleculares.

A primeira classificação dos TEs, é baseada no mecanismo de transposição e dividem os elementos em Classe I, chamados de retrotransposons, e os de Classe II, conhecidos como transposons de DNA ou simplesmente transposons (Finnegan, D. 1989). Wicker et al. (2007), dentro dos tipos principais, dividiu os elementos em superfamílias acrescentando a análise das TSDs (*target-site duplications*) e a presença ou ausência de longas repetições terminais nos retrotransposons. Ainda assim, grandes problemas são encontrados com os sistemas de classificação já propostos, diante disso, o avanço dos estudos de forma específica em cada grupo de elementos, se faz de extrema

importância para ampliar o conhecimento acerca dos TEs, principalmente compreendendo como ocorre a estruturação dos distintos grupos, consonante às sequências moleculares e ao critério filogenético dos elementos. Posteriormente a classificação de Wicker et al. (2007), Kapitonov e Jurka (2008) propôs uma classificação baseada na enzimologia, semelhança e relações das sequências, assim, definiram os elementos do tipo 1 como sendo transposons de DNA e os tipo 2 como sendo os retrotransposons.

De início, de modo a facilitar o direcionamento da classificação dos TEs, seguindo o sistema de Wicker et al (2007), os elementos são divididos em elementos de classe I ou classe II. Os elementos de classe I ou retrotransposons, necessitam de uma molécula de RNA intermediando a transposição que ocorre com a presença de uma enzima transcriptase reversa (RT) (Levin e Moran, 2011). Esses elementos conseguem realizar cópias de si mesmo e se inserirem em distintas regiões do genoma, esse mecanismo é popularmente conhecido como "copia e cola" e eles compõe uma larga proporção na maioria dos genomas (Chalopin et al., 2015). Os TEs da classe II (transposons de DNA) não necessitam de um intermediário de RNA para sua mobilização. Duas subclasses e quatro ordens são categorizadas conforme o mecanismo de transposição que pode ser do tipo "corta e cola" (*cute and paste*) ou que são os transposons rolling-circle DNA e transposons de DNA autosintetizante (Kidwell, 2002; Pritham e Feschotte, 2007; Bao et al., 2009).

As classes de TEs possuem subdivisões que agrupam elementos mais semelhantes em nível de estrutura molecular e mecanismo de transposição. Diante disso, os retrotransposons são subdivididos em dois tipos: (i) os elementos que possuem longas repetições terminais (*long terminal repeat* - LTRs), que são retroinseridos no genoma pela ação de uma enzima Integrase; (ii) e os elementos que não possuem as longas repetições terminais (não-LTR) e que são retroinseridos por uma Endonuclease (Wicker et al., 2007; Kapitonov e Jurka, 2008). Segundo Wicker et al., 2007, quatro ordens são conhecidas dentro dos elementos não-LTR, são elas: SINE (*short interspersed nuclear element*); Penelope; LINE (*long interspersed nuclear element*) e DIRS, que se diferem dos LTRs por apresentarem repetições terminais flanqueando uma região interna contendo genes (Capello et al., 1985; Goodwin e Poulter 2001).

Entre os transposons de DNA, a superfamília TIR possui a presença de uma transposase e são representados por nove superfamílias: Tc1-Mariner, hAT, Mutator, Merlin, Transib, P, PiggyBac, PIF-Harbiner e CACTA, em que se destacam os elementos Mariner-like (MLEs) por provavelmente serem os elementos com a maior distribuição nos genomas (Wicker, et al., 2007). Além disso, mais três superfamílias são reconhecidas

para os TEs de classe II, que são a Crypton, Helitron e Maverick (Wicker et al., 2007). As subdivisões internas apresentadas, variam conforme o sistema de classificação, Kapitonov e Jurka (2008) classifica os transposons de DNA em 17 superfamílias e os retrotransposons em 16 superfamílias de não-LTR e 7 superfamílias de LTR.

Presença e atuação dos elementos genéticos móveis nos genomas

Por mais que a grande maioria dos estudos, que verifica a presença dos TEs nos genomas, sejam baseadas na anotação dessas sequências, diversos trabalhos têm sido direcionados com o intuito de abranger os mecanismos de evolução molecular dos TEs no genoma dos diversos tipos de organismos. Trabalhos filogenéticos analisando as relações que ocorrem entre as sequências buscam compreender a evolução independente dos TEs que ocorre após a divergência das espécies, identificando, por exemplo, algumas características particulares à espécie em elementos que pertençam à mesma superfamília.

Além disso, estudos que avaliam as funções genômicas atribuídas aos TEs, também vêm ganhando espaço no meio científico, principalmente pelo fato de que se comprovou que os TEs são os responsáveis pelos genomas não serem estruturas estáticas, mas sim dinâmicas (Muotri et al., 2007). Isso ocorre graças aos estudos que comprovam que a movimentação desses elementos consegue influenciar no funcionamento dos genomas, induzindo alterações genômicas do tipo rearranjos, inserções, deleções, duplicações promovendo inovações genéticas que alteram a estabilidade do genoma (Bohne et al., 2008; Senft e Macfarlan, 2021).

Eventos distintos são responsáveis pelo surgimento de um TEs em um genoma, por exemplo, eventos de mutação, recombinação e transferência horizontal. Esses eventos fazem com que o TEs tenha um "ciclo de vida" composto por etapas dentro dos genomas dos organismos em que se encontram. O "ciclo de vida" se inicia com a etapa de invasão, em que ocorre a duplicação do elemento e a reinserção da nova cópia em outro local do genoma, ou seja, é o "nascimento" do TE no respectivo genoma. Para que ocorra essa transposição, o TE necessita de toda a maquinaria necessária para se transpor corretamente e o fato de conseguirem se estabelecer no genoma, está relacionado a fatores genéticos de forma que ocorra a interação do TE com o genoma do hospedeiro (Kidwell e Lisch, 2000; Rouzic et al., 2007). Posteriormente, ocorre a fase de maturação, onde há um aumento do número de cópias do elemento no genoma que é uma consequência direta da transposição. Por fim, ocorre a fase de senescência do TE,

pode ser através do mecanismo de inativação (mutações) e silenciamento (RNA de interferência, metilação, etc) dos elementos, em que ocorre a degeneração tanto das regiões terminais, quanto da região com as ORFs e domínios necessários para que os elementos sejam potencialmente ativos, causando a interrupção das unidades repetitivas (Kidwell e Lisch, 2000; Pinsker et al., 2001; Brookfield, J. 2005; Charlesworth et al., 1994; Kelkar et al., 2020). O "ciclo de vida" dos TEs é o motivo de encontramos sequências completas e potencialmente ativas nos genomas, bem como sequências degeneradas e que não são potencialmente ativas.

O fato de os TEs conseguirem se transpor no genoma hospedeiro, de apresentarem um "ciclo de vida", tendo de início a invasão do TE, a propagação e aumento do número de cópias e por fim a inativação e silenciamento, promove um aumento na variabilidade genética dos organismos em que estão inseridos. Existe uma teoria que prevê a extinção dos TEs em organismos que possuem reprodução assexuada (Irina et al., 2005; Park et al. 2021). O contrário ocorre nos organismos que se reproduzem sexuadamente, pois eles apresentam uma maior variabilidade genética, tendo os TEs como "matéria-prima", em diversos processos de recombinação dos genomas, sendo assim, os TEs são ativados em diferentes gerações atuando em questões evolutivas das espécies (Wright et al., 2001; Kent et al., 2017).

Durante a evolução, em determinados momentos, os TEs podem passar por processos que os fazem ter um crescimento no número de cópias de forma exagerada, chamamos isso de "boom", essas ondas de invasão dos TEs podem ocorrer ao acaso e inúmeras cópias são distribuídas nos genomas, sendo o mecanismo de transposição uma das principais causas de rearranjos cromossômicos e modulação do genoma das espécies (Biémont et al., 2006). Essas cópias podem se inserir em regiões gênicas atuando em vias de regulação gênica, ou em regiões não gênicas, contribuindo para a diversificação das espécies e aumento dos genomas, como por exemplo, no genoma de anfíbios (Hellsten et al., 2010). As cópias oriundas das ondas de invasão de TEs, que ocorrem durante a evolução do genoma, podem ser encontradas de maneira dispersa no genoma ou acumuladas em tandem em alguns cromossomos. Quando as inserções ocorrem em regiões gênicas e/ou regulatórias, pode ocorrer superexpressão ou silenciamento gênico (Lexa et al., 2018). Um exemplo de desordens genéticas oriundas das inserções de TEs nos genomas, é o elemento LINE (Long interspersed nuclear *element*), mais especificamente, LINE-1 ou L1, abundante no genoma de seres humanos, resultando em doenças genéticas, como, por exemplo, câncer, doenças autoimunes e neuronais, pois resultaram em alterações na natureza dinâmica do genoma (Hancks e

Kazazian, 2016). Devido a essa capacidade de transposição do LINE estar associada, principalmente, ao desenvolvimento do câncer, a inibição da transposição do elemento no genoma de seres humanos, tornou-se uma estratégia efetiva para o tratamento de diversos tipos de cânceres (Zhang et al. 2017).

A mobilização dos TEs pode resultar em "sítios alvo de duplicação" (*target site duplication* - TSD) que surgem no local onde ocorreu a inserção do elemento (Lavie et al., 2004). Vogt et al. (2014) verificou que esse mecanismo de mobilização dos TEs pode gerar deleções no genoma dos organismos, de maneira pontual, em alguns pares de bases ou até mesmo, grandes deleções gênicas, impedindo que o desenvolvimento do zigoto seja efetivo, sendo letal para os organismos (Rishishwar et al., 2015).

Devido ao caráter repetitivo dessas sequências no genoma e por muitas dessas cópias estarem dispersas em diferentes sítios cromossômicos, TEs podem favorecer eventos de recombinação homóloga não-alélica (Non-Allelic Homologous Recombination - NAHR). Eventos de NAHR podem gerar deleções, duplicações segmentais, inversões pericêntricas e paracêntricas (Hancks e Kazazian 2012; Klein e O'Neil, 2018) sendo as duas primeiras as mais comuns em NAHR intercromossômicos (Kazazian e Moran, 2017) e as inversões em em NAHR intracromossômicas (Beck et al., 2011), devido a presença de sequências repetidas em regiões distintas de um mesmo cromossomo. Portanto, as NAHR podem promover barreiras reprodutivas entre indivíduos de mesma população (Brown e O'Neill, 2010) devido as alterações cromossômicas estruturais, interferindo diretamente no comportamento da divisão meiótica das células reprodutivas.

Um exemplo envolvendo NAHR e TEs é o caso do isolamento reprodutivo entre os Neandertais e seres humanos modernos, contando com mais de 900 alterações na estrutura genômica desses organismos (Rogers, 2015). Rogers (2015) identificou diversas incompatibilidades, como por exemplo, rearranjos envolvendo o cromossomo 14 e o cromossomo 15, região responsável pelos receptores olfativo e também verificou pontos de rearranjos entre o cromossomo X e autossomos. Rogers (2015) sugere que os pontos envolvendo os TEs e NAHR podem atuar na homogeinização das sequências no genoma ou servirem como *hotspots* para a transposição e inserção de elementos. Por fim, mesmo com o silenciamento de expressão, os TEs ainda podem influenciar rearranjos cromossômicos pelo fato de favorecerem o mecanismo de recombinação ectópica (Petrov et al., 2011).

A barreira reprodutiva, também pode ocorrer com a formação de neocentrômeros junto a TEs (Schneider et al. 2016; Tolomeo et al. 2017). Henikoff et al. (2001) Henikoff et al. (2002) observou que as sequências presentes nos centrômeros são altamente

variáveis entre as espécies e que existe uma força seletiva atuando nas sequências de DNA do centrômero. Em espécies vegetais já foi apresentado retroelementos específicos atuando no centrômero de distintos organismos, como ocorre em milho com uma grande proporção do elemento CR2 (Bilinski et al., 2015). Em animais, a diversificação centromérica está diretamente relacionada a rearranjos cromossômicos envolvendo TEs (Carbone et al, 2014). Esses autores buscaram caracterizar o conteúdo de sequências repetitivas associadas às proteínas CENPs (*Centromere protein S*) nos centrômeros dos cromossomos na espécie de macaco Gibão. Eles encontraram elementos do tipo SINE, LINE e elementos LTRs nas regiões centroméricas, porém, os elementos SINEs foram os mais abundantes associados às CENPs.

Diante das alterações provocadas pela transposição dos TEs nos genomas, é de se esperar que esses elementos se mostrem potencialmente ativos ou então silenciados no genoma, devido ao acúmulo de mutações, por exemplo. Silenciados ou não nos genomas, caso presentes na linhagem gamética, eles serão transmitidos verticalmente seguindo os padrões naturais de segregação, ou seja, a transmissão é hereditária (McLaughlin e Malik, 2017; Dechaud et al., 2019).

Mecanismos epigenéticos também podem estar associados ao comportamento dos TEs nos genomas. O silenciamento dos TEs, relacionados a sua expressão, podem acontecer tanto no momento pré-transcrional quanto pós-transcricional. A influência epigenética pode atuar diretamente em mecanismos pré-transcricionais, por exemplo, na modificação da cromatina, por metilação do DNA (Kang et al., 2006), ou remodelação da cromatina (Kang et al., 2006) por meio de alterações covalentes nas caudas de histonas (Slotkin e Martienssen, 2007) impedindo que ocorra a transcrição do TE por impedir o acesso da maquinaria necessária para transcrever (Klein e O'Neil, 2018). Com relação ao mecanismo pós-transcricional, o silenciamento pode ocorrer por meio de RNA de interferência (iRNA), que é um mecanismo celular que funciona com a atuação de RNAs de fitas duplas (dsRNA) de forma a silenciar os genes (Horman et al., 2006). Este mecanismo já foi observado em *Caenorhabditis elegans* regulando a expressão do TE Tc1 por meio do RNA, denominado BCMVCP-dsRNA (Phillips et al., 2015).

Outros mecanismos de silenciamento envolvem RNAs associados a PIWI, chamados de PIWI RNA (piRNA), que são RNAs curtos não codificantes que agem através de um complexo de silenciamento induzido por piRNA (piRISC), podendo agir tanto na fase pré-transcricional, quanto na pós transcricional, além de atuar nas células germinativas dos organismos, silenciando TEs como forma de defesa do genoma (Siomi et al. 2011). As sequências de piRNAs são extremamente variadas, ou seja, existe uma

diversidade de sequências não codificantes que podem atuar regulando o silenciamento de RNA. Este silenciamento pode ocorrer envolvendo diversos processos, por exemplo, por proteínas dedo de zinco KRAB (KZFP) e seu corepressor, TRIM28, que se associam a regiões específicas atuando na regulação gênica da expressão das proteínas (Wolf et al. 2015).

De acordo com Bourgeois et al. (2019), os TEs podem ser silenciados por um mecanismo de seleção com base em seu potencial de ligação com um alelo vantajoso, ou seja, caso este elemento se encontre muito próximo de algum alelo que seja essencial ao organismo, os TEs podem passar por um processo de silenciamento a fim de se preservar o genoma do organismo em que estão inseridos. Caso não ocorra o silenciamento, os TEs podem ser "domesticados" favorecendo a criação de novos genes (Enriquez-Gasca et al., 2020), como acontece, por exemplo, em mamíferos, em que os genes favorecem o desenvolvimento da placenta (Gilbert et al., 2018).

Elementos genéticos móveis nos genomas de Anuros e a ordem DIRS

Em anuros, embora ainda seja muito restrito o número de genomas sequenciados, montados e disponíveis para análises, verifica-se uma diversidade expressiva de elementos nos genomas (Hellsten et al., 2010; Chalopin et al., 2015; Keinath et al., 2015; Sotero-Caio et al., 2017). Estima-se que aproximadamente 33% do genoma de *Xenopus tropicalis* é composto por TEs sendo cerca de 75% composto por elementos de classe II (Hellsten et al., 2010; Chalopin et al. 2015). O contrário ocorre no genoma de *Nanorama parkeri* em que os retrotransposons (especialmente os LTRs – *long terminal repeats*) são os elementos mais abundantes (Sun, et al., 2015; Sotero-Caio, 2017). Além desses genomas, *Rhinella marina* apresenta um acúmulo de elemento retrotransposon do tipo LINE (Edwards et al., 2018).

Apesar da maioria dos estudos com relação aos retrotransposons serem pautados nos elementos LTR e não-LTR (Goodwin e Poulter, 2001; Sun et al., 2011), os elementos da ordem DIRS (*Dictyostelium discoideum transposon DIRS-1*), descritos pela primeira vez no genoma de *Dictyostelium discoideum* (Cappello et al., 1985), vêm ganhando especial atenção recentemente, especialmente por que estão sendo classificados recentemente como um grupo independente dos LTRs (Goodwin e Poulter, 2004; Poulter e Goodwin, 2005). O fato desses elementos já terem sido considerados dentro dos elementos LTR, pode ter contribuído para quantificações errôneas do conteúdo de

retrotransposons em diversos genomas e por consequência subestimado a presença desses elementos.

Esses elementos da ordem DIRS podem ser diferenciados dos demais retrotransposons por codificarem um gene para a enzima Tirosina Recombinase (YR) em vez de uma integrase ou endonuclease e apresentam, de modo geral, três ORFs: ORF1 apresenta um domínio para Gag; ORF2 que codifica uma transcriptase reversa (RT) e uma RNAse H (RH); e a ORF3 que frequentemente aparece sobreposta com a ORF2, codifica uma YR (Cappello et al., 1985). Alguns elementos DIRS apresentam uma região codificadora na ORF2 para uma metiltransferase (MT) que não se sabe ao certo sua função específica nesses elementos (Goodwin e Poulter, 2004; Poulter e Goodwin, 2005). Com relação as repetições que aparecem nas extremidades desses elementos, elas possuem uma estrutura diversificada conforme cada uma das superfamílias da ordem DIRS e podem ser de dois tipos. O primeiro tipo caracterizado pela presença de repetições terminais invertidas (terminal inverted repeats -ITRs) presentes nas porções 5' e 3' e adicionalmente observa-se uma região complementar interna (internal complementary region – ICR), que possui uma sequência nucleotídica complementar no início do ITR da porção 5' (esquerdo) e no final da porção 3' (direito) do ITR. O segundo tipo de repetição terminal apresenta as chamadas repetições diretas divididas (split direct repeats - SDRs) sendo que na porção inicial 5' apresenta a repetição do tipo A1, seguida por B1 e na porção final 3' B2, seguida por A2, sendo A1 e A2 idênticas, bem como B1 e B2 (Cappelo et al., 1985; Goodwin et al. 2004; Poulter e Butler 2015).

Ribeiro et al. (2019) utilizando o genoma de organismos kinetoplastídeos, que são protistas flagelados, como por exemplo, *Trypanosoma e Leishmania*, baseados no critério filogenético utilizando a sequência da RT, RH e YR dos elementos que codificam uma YR, classificou os elementos desses organismos em quatro superfamílias, são elas: *DIRS-like*. *Ngaro-like, PAT-like e VIPER-like*. Os elementos *DIRS-like* são caracterizados pela presença repetições do tipo ITRs e possuem as três ORFs com a característica peculiar de que em algumas famílias ocorre a presença da metiltransferase na ORF 2 (Goodwin et al., 2004; Poulter e Butler, 2015). A superfamília *Ngaro-like* apresenta as repetições terminais do tipo SDRs e apresentam como particularidade de algumas famílias, após o domínio da YR, uma ORF adicional contendo um domínio da hidrolase (Hydro – SGNH) (Goodwin et al. 2004; Poulter e Butler 2015). A superfamília *PAT-like* apresenta uma singularidade de serem internamente semelhantes em relação as ORFs e os domínios conservados aos *DIRS-like* e possuírem as repetições terminais do tipo SDRs, ou seja, semelhantes aos *Ngaro-like* (Poulter e Goodwin 2005; Ribeiro et al. 2019). Por fim, os

elementos da superfamília *VIPER-like* também possuem SDRs nas extremidades e são um grupo de retrotransposons que até o momento parece estar restrito aos kinetoplastídeos (Ribeiro et al., 2019). Os elementos DIRS já foram encontrados em representantes da ordem Anura, como por exemplo, nos genomas de *X. tropicalis* e *X. laevis* (Goodwin et al. 2004; Hellsten et al. 2010; Poulter e Butler 2015), *Rhinella marina* (Ludwig et al., 2021), *Ololygon centralis* (Castro, A., 2018), *Nanorana parkeri* (Castro, A., 2018), porém, esses autores apenas reportaram esses elementos na anotação dos genomas, sem realizar um estudo aprofundado sobre sua diversidade e estrutura molecular (Figura 1).



Figura 1: Representação esquemática dos elementos da ordem DIRS baseada nos presentes no genoma de *X. tropicalis e X. laevis* de Gazolla et al. (2022) (a) *DIRS-like,* (b) *NGARO-like* e (c) *PAT-like.* Todos os elementos são representantes da ordem DIRS de retrotransposons.

Diante disso, trabalhos que combinem a caracterização de TEs seguidos de mapeamento cromossômico, permitindo a detecção dos locais cromossômicos de acúmulo desses elementos, desperta interesse na compreensão dos TEs junto ao estudo da evolução do genoma e por consequência, evolução dos cariótipos entre as espécies intimamente relacionadas.

Família Pipidae: um interessante grupo para estudo da evolução cromossômica e genômica

A ordem Anura reúne espécies de anfíbios popularmente conhecidos como sapos, pererecas e rãs, e abarcam 7424 espécies das 8408 descritas (Frost, 2022. Acesso em Fevereiro de 2022). A família Pipidae reúne atualmente 41 espécies, divididas em duas subfamílias, sendo Dactylethrinae contendo 35 espécies, enquanto Pipinae é monotípica, somente com as Pipas contendo 7 espécies (Boulenger, A. 1882). Além disso, quatro gêneros são propostos para essa família, sendo, o gênero *Hymenochirus* Boulenger, 1896 (4 espécies), gênero *Pipa* Laurenti, 1768 (7 espécies), gênero *Pseudhymenochirus* Chabanaud, 1920 (1 espécies) e gênero *Xenopus* Wagler, 1827 (29 espécies) (Cannatella e Trueb, 2008).

O genoma de *X. tropicalis* foi o primeiro a ser sequenciado na ordem Anura isso faz com que ele seja usado como modelo de referência em estudos genômicos comparativos e evolução cromossômica (Hellsten et al., 2010). Além disso, o processo de especiação entre *X. tropicalis e X. laevis* que ocorreu há pelo menos 48 milhões de anos (Session et al. 2016), chama a atenção de estudos acerca do genoma de *X. laevis* devido a origem alotetraplóide (cerca de 17-18 milhões de anos), possuindo atualmente dois subgenomas, denominados L e S que se originaram de dois progenitores diplóides separados a aproximadamente 34 milhões de anos. Devido o evento de alotetraploidização, ambos os subgenomas sofreram diversificação (Session et al. 2016). Portanto, estudos comparativos entre esses genomas e estudos direcionados ao entendimento específico de cada um, são de grande relevância para a compreensão da evolução do grupo.

Recentemente o genoma de *P. carvalhoi* foi sequenciado por nosso grupo de pesquisa (Laboratório de Citogenética Evolutiva e Conservação Animal, da Universidade Federal do Paraná) utilizando a plataforma Illumina HiSeq-400, cobertura 46x (GAC021901965.1). O genoma foi montado com Velvet 1.2.10 (Zerbino e Birney, 2008) pelo método *de novo* com um parâmetro de valor de comprimento de 105 de hash (kmer) e atualmente está a nível de contigs (2,4 Gpb, 5.029.551 contigs).

Do ponto de vista cromossômico, as três espécies apresentam interessantes questões. Roco et al. (2021) pela técnica de GISH (*genomic in situ hybridization*), comparando o genoma das duas espécies, identificou, no genoma de *X. tropicalis*, sinais positivos em alguns cromossomos, como por exemplo, nas extremidades de todos os cromossomos metafásicos, além disso, marcações foram observadas no braço curto do cromossomo 3 e na constrição secundária do cromossomo 9. Não foi observada diferenças relacionadas ao sexo. *X. laevis* mostrou um padrão diferente do observado em *X. tropicalis*, os sinais observados nas extremidades dos cromossomos 3L e no centrômero do 6S. Utilizando a sonda de um genoma no outro, em *X. tropicalis* apareceram marcações no final de todos os cromossomos do cariótipo e em *X. laevis*, além das marcações nas extremidades de todos os cromossomos, duas marcações específicas apareceram nos cromossomos 3S e 3L.

Comparações cromossômicas realizadas por Zattera et al. (2020) entre *X. tropicalis* e *P. carvalhoi,* detectaram marcações da sonda de histona H3 nos cromossomos 5 e 6 em *P. carvalhoi, X. tropicalis e X. laevis*, e nos pares 8 e 9 nos cromossomos de *P. carvalhoi e X. tropicalis.* Além disso, marcações utilizando sondas de microssatélites (CAG)10, (CGC)10 e (GATA)10, mostraram padrões compartilhados entre *X. tropicalis e P. carvalhoi.*

Diante dos expostos, propomos abordagens de bioinformática e citogenética combinadas, para avaliar os elementos de transposição em nível genômico e cromossômico as espécies *X. tropicalis, X. laevis* e *Pipa carvalhoi.* Como dito anteriormente, é fato a presença de elementos transponíveis atuando na reorganização dos genomas em pontos de rearranjos cromossômicos, e os resultados de ambas as técnicas se complementando, podem contribuir para entender a diversificação cariotípica entre essas espécies filogeneticamente relacionadas.

OBJETIVOS

Objetivos gerais

Investigar, por meio de análises genômicas e citogenética, os retrotransposons da ordem DIRS no genoma de *X. tropicalis*, *X. laevis* e *Pipa carvalhoi.*

Objetivos específicos

Analisar a estrutura molecular, diversidade genômica e padrões evolutivos de cópias de retrotransposons da superfamília *DIRS-like* e *Ngaro-like* presentes no genoma de *X. tropicalis* e *X. laevis;*

Avaliar a distribuição cromossômicas cópias de sequências homólogas aos retrotransposons da superfamília *DIRS-like* e *Ngaro-like* nos cariótipos de *X. tropicalis* e *P. carvalhoi*;

Buscar por regiões cromossomo-específicas entre *X. tropicalis* e *P. carvalhoi* por meio de ensaios de CGH (*Comparative Genomic Hybridization*)

CAPÍTULO 1

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

Artigo publicado no periódico G3: Genes, Genome, Genetics

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[§]; DanielPacheco Bruschi[†]

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

- 3 Laboratório de Ciências e Tecnologias Aplicadas em Saúde (LaCTAS), Instituto Carlos Chagas – Fiocruz-PR, Curitiba, PR, Brazil.
- 4 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 = 20and 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' noncoding terminal regions, preserved open reading frames, 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 both subgenomes of X. laevis. More recent amplification waves are seen in X. tropicalis (less than 3.2 mya) and X. laevis (around 10 mya) corroborating with transcriptional activity evidence. All DIRS-like families were found in both X. laevis subgenomes, while a few were most represented in the L subgenome. Ngaro-like elements presented less diversity and quantity in X. tropicalis and X. laevis genomes, although potentially active copies were found in both species and this is consistent with a recent amplification wave seen in the evolutionary landscape. Our findings highlight a differential diversity-level and evolutionary dynamics of the YR retrotransposons in 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.* 2015; 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 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 [long terminal repeats (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 noncoding 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 variations 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 *Xenopus 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 where the two subgenomes (called S and L) are identified (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* 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 Rebpase 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 ITRs, ICRs, and 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 available for each sequence 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 (aa) 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_017654675.1) genomes. The first 10 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 aa sequences, including the following sequences: (1) the consensus sequences of the DIRS families of *X. laevis* and *X. tropicalis* recovered from Rebpase; (2) 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 (3) 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 (Rebpase—*Saccoglossus kowalevskii*), and *PAT* (Q26106—*Panagrellus redivivus*), and *DIRS-like*—DIRS-1_Acar (Rebpase—*Anolis carolinensis*) and DIRS-5_CBP (Rebpase—*Chrysemys picta bellii*).

The sequences were aligned using the PSI-coffee tool (Notredame *et al.* 2000), with Genedoc 2.7 (Nicholas and Nicholas 1997) being used for sequence manipulation and editing. Most of the RT sequences contained around 120 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 aa substitution model. A distance tree was constructed using the Neighbor-joining method with JTT + G model and bootstrap test with 1000 replicates. A phylogenetic tree was also reconstructed using Bayesian inference, run in MrBayes 3.2.6 (Ronquist *et al.* 2012) based on the LG + G model. The Markov Chain Monte Carlo was run for 10,000,000 generations, sampled every 1000 generations, with 25% of the initial results being discarded as burn-in. The final trees were 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 evolutionary 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 libraries of each species. In the case of the *Ngaro-like* elements of *X. laevis*, as the families were not available in the Rebpase, 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 and Nicholas 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). For *X. laevis*, the subgenomes S and L were screened separately. 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 evolutionary landscape were created using the Python Matplotlib-v3.3.2 (Hunter 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 (Bowes *et al.* 2010) 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 Rebpase for *X. tropicalis* and 20 for *X. laevis* (Supplementary Table S1). In the Repbase classification, all YR-containing elements are classified as superfamily DIRS, a final group within the LTR retrotransposons group (Kapitonov and Jurka 2008). Mainly concerning the YR elements, the Wicker *et al.* (2007) classification is more detailed, separating them into a distinct group of retrotransposons (order DIRS) and discriminating the clear distinct subgroups of DIRS into three superfamilies and more recently separation into four superfamilies has been suggested (Ribeiro *et al.* 2019). We thus previously assume that these DIRS families available in the Repbase could belong to any of the DIRS superfamilies, then our analyses indicate they belong only to *DIRS-like* and *Ngaro-like*.

The evolutionary trees based on the RT domain of all the elements present similar topologies and recovered two well-supported clades (Figure 1 and Supplementary Figure S1) in which all the *Xenopus* sequences grouped in either (1) a *DIRS-like* or (2) a *Ngaro-like* group. The two *PAT-like* sequences were not grouped as a monophyletic group.
The divergence between these two groups of sequences is clear (Figure 1). We recognized only two of the 95 *Xenopus* Rebpase 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 Repbase 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 Supplementary Tables S1 and S2, for more details).

The ORF1 encodes a gag-like protein and a LAP2alpha domain (~650 aa) was predicted in all 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 38 of the families evaluated here. The ORF3 encodes the YR protein with a conserved DNA_BRE_C domain of around 584 aa (Supplementary Table S2).

The *DIRS-like* elements have 5' and 3' ITRs and an ICR region (Figure 2A), and this pattern of repeats was found in almost all the *Xenopus DIRS-like* families evaluated here. The ITRs have ~120 bp and present a few nucleotide substitutions or indels between the left ITR (IITR) and the right ITR (rITR) (Figure 2A). The ICR is composed of two short sequences (IICR and rICR), which are complementary to the 5' (ITer) 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 (Supplementary Table S2). In *X. laevis*, 13 of the 20 families present premature interruptions in the ORFs or incomplete repeats (Supplementary Table S2), which

indicates a high level of degeneration in these families. Concerning the EST data, we observed that most families present transcripts (55 families from *X. tropicalis* and 12 families from *X. laevis*) (Supplementary Table S1). 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 family diversity of the *DIRS-like* clade in *Xenopus* (Figure 1). The diagnostic *DIRS-like* sequences from Sauropsids (DIRS-1_ACar and DIRS-5_CBP) were recovered as a basal branch, which indicates that most of *DIRS-like* families' diversity was originated after the separation of amniotes and amphibians, although in the BA tree (Supplementary Figure S1), these sequences were located inside the *Xenopus DIRS-like* clade.

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-14_XL + DIRS-17_XL, DIRS-2_XL and DIRS-50_XT, 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 evolutionary landscape profile observed in each genome further reinforces this conclusion (Figure 3). For *X. laevis*, in both subgenomes, it is possible to observe two broad ancient waves of amplification [60–80 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 younger copies in *X. laevis* genome are found in smaller proportions than in *X. tropicalis* and a small peak of recent amplification is seen around 10 mya. Although the diversity of the *DIRS-like* families is much lower in *X. laevis*, they make up a larger proportion of the genome (1.03%) than in *X. tropicalis* (0.5%) with similar proportion in both subgenomes (0.45% for S subgenome and 0.58% for L subgenome). The families DIRS-2_XL, DIRS-3_XL, DIRS-10_XL, DIRS-12_XL, DIRS-13_XL, and DIRS-14_XL present a slightly higher proportion of mapping (around 60%) in the L genome (Supplementary Figure S2).

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 1–7 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. The nearest sequences to the query are the copies 1, 2, 4, and 5 (copy 3 was not included in the tree) forming a clade with no known family, suggesting that these sequences may be copies of a *X. laevis* family that has yet to be established. Copies 6–10 grouped with DIRS-12_XL, which is closely related to DIRS-1_XL. We found conserved copies 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 S2). 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 A1 and A2 repeats of 236 bp and B1 and B2 repeats of 152 bp, both being 100% identical. No SDRs were found in the DIRS-54_XT. Although none of the *X. laevis* Rebpase 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 (Supplementary Table S1).

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 S2), 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*, we observed that 13 homologous copies of DIRS-53_XT/DIRS-54_XT are preserved, while some include a complete or partial loss of the 5' and/or 3' SDRs or

broken ORF (Supplementary Table S2). Copies 2, 3, 4, 7, 8, and 9 recovered with DIRS-53_XL and copies 1, 2, 3, 4, 5, 8 and 10 recovered with DIRS-54_XL are 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). Clearly, DIRS-53_XT and DIRS-54_XT are very divergent (~65% over only 600 bp of good alignment) and form two major groups of *Ngaro-like* sequences; however, some of the relationships were not supported by the bootstrap values and the Bayesian tree (Supplementary Figure S1) has shown a distinct topology, grouping some of the copies recovered with DIRS-54_XT in both species as a basal clade together with DIRS-53_XT. In both trees, the structuring of subclades with species-specific groups suggests the activity of these families after the separation of *X. tropicalis and X. laevis*. While the SDRs are not present in the consensus canonical copy of DIRS-54_XT, they can be observed in copies 5 and 10, which indicates that this family is also potentially active in *X. tropicalis*.

Similar to what we observed for *DIRS-like*, *Ngaro-like* elements density in *X. laevis* (0.091%) is approximately double of the density seen for *X. tropicalis* (0.044%) and these proportions are 10 times smaller than observed for *DIRS-like* in both genomes. The *Ngaro-like* evolutionary landscape profile (Figure 3) shows a recent amplification signal (less than 3.2 mya) in *X. tropicalis* genome and *X. laevis* subgenomes. Additionally, in both species, it is possible to observe a very ancient amplification wave, dating from 120 to 145 mya, that is more prominent in *X. laevis* 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' noncoding 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; Poulter and Goodwin 2015). Our detailed analysis of the elements corroborates what has been described for these superfamilies concerning the ORFs and domains (Goodwin and Poulter 2001; Poulter and Goodwin 2005). We found the three main ORFs for both

superfamilies and a MT domain or a hydrolase domain was found in the *DIRS-like* and *Ngaro-like* families, respectively.

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 is probably 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 for 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.

Based on the phylogenetic criteria and molecular structure, we found that both superfamilies are present in both genomes. The presence of *DIRS-like* and *Ngaro-like* was already reported for *X. tropicalis* (Hellsten *et al.* 2010). On the other hand, for *X. laevis* (Goodwin and Poulter 2004), the information was not clear and no *Ngaro-like* family was deposited in the Repbase dataset version that we analyzed. 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. This richness of *DIRS-like* elements in *Xenopus* is also higher than that found for the fish *D. rerio* (12 families) or the reptile *A. carolinensis*, which has 42 families (Piednoël *et al.* 2011). *DIRS-like* and *Ngaro-like* elements are widely distributed in a number of metazoan groups, including fish, amphibians, reptiles, and some fungus (Ruiz-Perez *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.

Several sequences of both superfamilies are structurally complete in *X. tropicalis* and *X. laevis*. Thus, it is relevant to compare the evolutionary landscape pattern of these superfamilies in both genomes since these species have undergone distinct evolutionary processes after the split. The speciation of *X. tropicalis* and the *X. laevis* ancestor at around 48 mya (Session *et al.* 2016; Figure 3). 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). 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 shows different patterns in *X. tropicalis* and *X. laevis* genomes.

For *DIRS-like*, looking at the *X. laevis* subgenomes graphs, it is possible to suggest that two ancient waves of amplification (60-80 mya and 110-135 mya) have occurred indicating the long-time of persistence of these elements. Considering the time of the species splitting, these waves probably occurred in the ancestor of both species, while these 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. All families were mapped in both X. laevis subgenomes and for a few families, the genomic density was slightly higher in the subgenome L. It is suggested that 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 (Session et al. 2016). The small signature of a recent wave of amplification in both subgenomes occurred after the allotetraploid origin of X. laevis resulting in a small proportion of younger copies that have been maintenance preserved for a long-time in X. laevis genome. In X. tropicalis, the DIRS-like 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 ours searches in the EST libraries of both species (Supplementary Table S1). Recent burst events in both species may also have contributed to the diversification of families in each genome, particularly in X. tropicalis, which is consistent with the strong species-specific grouping seen in the sequence tree.

The evolutionary landscape pattern and the phylogenetic trees 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 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. The intragenomic behavior of the TEs 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.

So far, this is the most comprehensive work of DIRS retrotransposons in Amphibia being a starting point to guide research on the evolution and functionalities of these retrotransposons in other anuran genomes and in this way solve gaps in evolutionary history not only in the way they behaved during evolution but also in how they influenced their genomes.

Data availability

Supplementary files available at Table S1 contain description of basic features of the DIRS families of *Xenopus tropicalis* (XT) and *X. laevis* (XL) deposited in Rebpase. 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. Supplementary Table S2 contains summary analyzes of the

copies retrieved of the *X. tropicalis* and *X. laevis* genomes. Supplementary Figure S1 contains sequence tree produced by Bayesian inference, based on the amino acid sequences of the RT domain. The matrix was composed of the sequences of *X. tropicalis* and *X. laevis* DIRS elements obtained from the Rebpase database, the copies retrieved from both genomes and diagnostic sequences from each DIRS superfamily and Supplementary Figure S2 contains the proportion of all *X. laevis DIRS-like* families in the subgenomes S and L. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. Supplementary material available at figshare: https://doi.org/10.25387/g3.15105336.

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 (<u>www.xenbase.org</u>) 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 Rebpase. 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	Superfamil y	Preserved ORFs	Conserved structure of repeats	EST
DIRS-1_XT	5556 bp	DIRS-like	Yes	No	Yes
DIRS-2_XT	5992 bp	DIRS-like	No, 1 and 3 not preserved	Yes	Yes

DIRS- 2A_XT	6224 bp	DIRS-like	Yes	Yes	Yes
DIRS- 3A_XT	5809 bp	DIRS-like	Yes	Yes	Yes
DIRS-3_XT	5870 bp	DIRS-like	Yes	Yes	Yes
DIRS- 4A_XT	5718 bp	DIRS-like	Yes	Yes	Yes
DIRS- 4B_XT	5380 bp	DIRS-like	Yes	No	No
DIRS-4_XT	5616 bp	DIRS-like	Yes	Yes	Yes
DIRS-5_XT	5682 bp	DIRS-like	Yes	Yes	Yes
DIRS- 6A_XT	5710 bp	DIRS-like	Yes	No	Yes
DIRS-6_XT	5814 bp	DIRS-like	Yes	Yes	Yes
DIRS-7_XT	5503 bp	DIRS-like	Yes	Yes	Yes
DIRS- 8A_XT	5835 bp	DIRS-like	Yes	Yes	Yes
DIRS-8_XT	5898 bp	DIRS-like	Yes	Yes	Yes
DIRS- 9A_XT	5447 bp	DIRS-like	Yes	No	Yes
DIRS-9_XT	5386 bp	DIRS-like	Yes	Yes	Yes
DIRS- 10_XT	5129 bp	DIRS-like	No, 3 not preserved	Yes	No
DIRS- 11 XT	5426 bp	DIRS-like	Yes	Yes	Yes
DIRS- 12_XT	5162 bp	DIRS-like	Yes	Yes	Yes
DIRS-	5629 bp	DIRS-like	Yes	Yes	Yes
13A_XT	•				
DIRS-	5825 bp	DIRS-like	Yes	Yes	No
13B_XT					
DIRS-	5716 bp	DIRS-like	Yes	Yes	Yes
13C_XT					
DIRS- 13_XT	5777 bp	DIRS-like	Yes	Yes	Yes
DIRS- 14_XT	5878 bp	DIRS-like	Yes	Yes	Yes
DIRS- 15_XT	5270 bp	DIRS-like	No, 3 not preserved	Yes	No

DIRS- 16_XT	5346 bp	DIRS-like	Yes	Yes	Yes
DIRS-	5302 bp	DIRS-like	Yes	Yes	No
17A_XT					
DIRS- 17_XT	5368 bp	DIRS-like	Yes	Yes	Yes
DIRS- 18 XT	5685 bp	DIRS-like	Yes	Yes	Yes
DIRS- 19 XT	5643 bp	DIRS-like	Yes	Yes	Yes
DIRS- 20 XT	5075 bp	DIRS-like	Yes	Yes	No
DIRS-	5716 bp	DIRS-like	Yes	Yes	Yes
21A_XT					
DIRS21_XT	5083 bp	DIRS-like	Yes	No	No
DIRS- 22A XT	5059 bp/ 5298 bp*	DIRS-like	No, 2 not preserved	No	No
– DIRS- 22_XT	5608 bp	DIRS-like	No 3, not preserved	Yes	Yes
DIRS- 23 XT	5277 bp	DIRS-like	Yes	Yes	Yes
DIRS- 24_XT	5211 bp	DIRS-like	No, 3 not preserved	Yes	No
DIRS-	5316 bp	DIRS-like	Yes	No	No
DIRS-25_XT	5504 bp	DIRS-like	Yes	Yes	Yes
DIRS-	5370 bp	DIRS-like	Yes	Yes	No
26A_XT					
DIRS- 26_XT	5415 bp	DIRS-like	Yes	Yes	Yes
DIRS-	5061 bp	DIRS-like	No, 3 not preserved	No	Yes
27A_XT			L		
DIRS- 27_XT	5611 bp	DIRS-like	No, 2 not preserved	No	Yes
DIRS-	5644 bp	DIRS-like	Yes	Yes	Yes

DIRS- 29_XT	5665 bp	DIRS-like	No, 1 not preserved	Yes	Yes
DIRS- 30_XT	5781 bp	DIRS-like	Yes	Yes	Yes
DIRS- 31_XT	5691 bp	DIRS-like	Yes	Yes	Yes
DIRS- 32_XT	5187 bp	DIRS-like	Yes	Yes	No
DIRS- 33_XT	4146 bp	DIRS-like	No, 2 not preserved	Yes	Yes
DIRS- 34A_XT	5490 bp	DIRS-like	Yes	No	Yes
DIRS- 34_XT	2880 bp/4728 bp*	DIRS-like	No, 2 not preserved	No	No
DIRS- 35_XT	5900 bp	DIRS-like	No, 2 not preserved	No	Yes
DIRS- 36A_XT	5595 bp	DIRS-like	No, 2 not preserved	No	No
DIRS- 36B_XT	5562 bp	DIRS-like	Yes	No	Yes
DIRS- 36_XT	5415 bp	DIRS-like	Yes	No	Yes
DIRS- 37A_XT	5123 bp/5602 bp*	DIRS-like	No, 3 not preserved	Yes	Yes
DIRS- 37_XT	5643 bp	DIRS-like	Yes	Yes	Yes
DIRS-38_XT	5221 bp	DIRS-like	Yes	Yes	Yes
DIRS- 39_XT	5484 bp	DIRS-like	No, 3 not preserved	Yes	Yes
DIRS- 40_XT	5448 bp	DIRS-like	Yes	Yes	Yes
DIRS- 41A_XT	2952 bp/5431 bp*	DIRS-like	No, 3 not preserved	No	No
DIRS- 41_XT	5212 bp	DIRS-like	No, 3 not preserved	Yes	No

DIRS- 42_XT	5710 bp	DIRS-like	No, 2 not preserved	No	Yes
DIRS- 43_XT	5156 bp	DIRS-like	No, 2 not preserved	Yes	Yes
DIRS- 44_XT	4762 bp	DIRS-like	Yes	Yes	Yes
DIRS- 45_XT	5054 bp	DIRS-like	Yes	No	Yes
DIRS- 46_XT	5239 bp	DIRS-like	No, 3 not preserved	Yes	No
DIRS- 47_XT	3912 bp/4490 bp*	DIRS-like	No, 3 not preserved	No	No
DIRS- 48_XT	5153 bp	DIRS-like	No, 2 not preserved	Yes	No
DIRS- 49_XT	5152 bp	DIRS-like	No, 2 not preserved	No	Yes
DIRS- 50_XT	5778 bp	DIRS-like	No, 3 not preserved	No	Yes
DIRS- 51_XT	4911 bp	DIRS-like	No, 1 and 2 not preserved	No	Yes
DIRS- 52_XT	5533 bp	DIRS-like	Yes	Yes	No
DIRS- 53_XT	5457 bp	Ngaro-like	Yes	Yes	Yes
DIRS- 54_XT	5141 bp	Ngaro-like	Yes	No	Yes
DIRS-1_XL	5291 bp	DIRS-like	Yes	Yes	No
DIRS-2_XL	5742 bp	DIRS-like	No, 1 and 3	Yes	Yes
			not preserved		
DIRS-3_XL	5387 bp	DIRS-like	No, 1 and 2 not preserved	Yes	No
DIRS- 4B_XL	5417 bp	DIRS-like	No, 2 not preserved	Yes	Yes

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

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

Figure legends:

Fig. 1 Sequence tree produced by neighbor-joining method (JTT + G), 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 Rebpase database, the copies retrieved from both genomes and diagnostic sequences from each DIRS superfamily. The bootstrap values higher than 50 are indicated at the branches. The * sign near the nodes indicates the clade was supported with posterior probability higher than 80 in the Bayesian tree. Sequences of different superfamilies are highlighted in different colors and shades of each color also distinguish the sequences of *X. tropicalis* (XT) and *X. laevis* (XL).

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 Rebpase consensus sequences of the DIRS-37_XT, that contain three ORFs, conserved domains (gag, RT/RH/MT, and YR) and noncoding 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 Rebpase 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. The evolutionary events as estimated by Session *et al.* (2016) are shown: the speciation of *X. tropicalis* and the ancestor of *X. laevis* at 48 mya, the speciation of the L and S progenitors of *X. laevis* at 34 mya, and their hybridization around 17–18 mya. The graphs show the divergence of *DIRS-like* (above) and *Ngaro-like* (below) copies mapped in the genomes of

X. tropicalis and *X. laevis* (S and L subgenomes) 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).

LITERATURE CITED

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 RecombinaseEncoding 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 TcDirs1 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) Rebpase 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 Rebpase. Nat Rev Genet, v. 9, n. 5, p. 411–2.

Kidwell M G, Lisch D R (2001). Perspective: Transposable Elements, Parasitic Dna, AndGenomeEvolution.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 Rebpase: RebpaseSubmitter 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: <u>http://www.nrbsc.org/gfx/genedoc/gdsrc.htm</u>.

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, Higuet 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/s13100-0190175-2 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 Systematic Across а Large Model Space. Biology, 61(3), 539-542. doi:10.1093/sysbio/sys029

Ruiz-Perez VL, Murillo FJ, Torres-Martinez S. (1996) Prt1 an unusual retrotransposonlike sequence in the fungus Phycomyces blakesleeanus. 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

Figure 1:

Tree scale: 1

DID DIRS-3 DIRS-34A DIRS-26 X DIRS-26A) DIRS-27 XT DIRS-39 XT D37 DIRS-46 XT DIRS DIRS-9 XT DIRS-5 DIRS-9A XT 94 100 -19 XT TCAX C2X **JIRS-**13 X DIRS



Figure 2:



CAPÍTULO 2 – Resultados

ESTUDO COMPARATIVO DAS FAMÍLIAS *DIRS-like* E *NGARO-like* NO GENOMA DE *Pipa carvalhoi* E *Xenopus tropicalis* COM EVIDÊNCIAS DE REGIÕES REPETITIVAS ESPÉCIE-ESPECÍFICAS EM *X. tropicalis*

RESUMO

Os genomas dos anuros apresentam uma diversidade de elementos de transposição que em sua grande maioria são simplesmente quantificados no genoma durante sua montagem e anotação, sendo poucos os estudos voltados para a sua estrutura molecular e distribuição em nível cromossômico. Neste trabalho investigamos os retrotransposons pertencentes a ordem DIRS no genoma de Pipa carvalhoi. Além disso, realizamos um estudo citogenético envolvendo as superfamílias DIRS-like e Ngaro-like em P. carvalhoi e X. tropicalis, bem como o CGH de ambas as espécies a fim de identificar possíveis regiões enriquecidas diferencialmente com DNA repetitivo (regiões espécie-específicas). Ambas as espécies apresentam um cariótipo de 2n=20 cromossomos e divergiram a aproximadamente 136 milhões de anos. Nosso estudo investigou a ocorrência e características moleculares dos elementos da ordem DIRS em P. carvalhoi. Nós conseguimos recuperar fragmentos de elementos neste genoma, porém, devido a fragmentação na montagem deste genoma, não conseguimos identificar nenhum elemento completo ou potencialmente ativo. O mapeamento cromossômico por FISH permitiu identificar em X. tropicalis regiões enriquecidas com os elementos DIRS-like, principalmente no braço curto do cromossomo 3, e de Ngaro-like, na constricção secundária do braço longo do par 7. Nos experimentos de Hibridação Genômica Comparativa (CGH) identificamos regiões espécie-específicas em ambos os organismos, sendo em P. carvalhoi, a região centromérica dos pares 1, 3 e 10; e em X. tropicalis, o braço curto do par 3, constrição secundária do braço longo do par 7 e região subterminal do braço curto do par 10. Além disso, as marcações cromossomos específicas de X. tropicalis, coincidiram com as marcações de TEs acumuladas no cromossomo 3 com DIRS-37 XT e no cromossomo 7, com Ngaro-53 XT. Já no genoma de P. carvalhoi não foi possível de detectarmos a composição do DNA repetitivo nas regiões específicas. Nossos achados reforçam a hipótese de que apesar de ambos os genomas apresentarem TEs de mesma ordem, eles possuem uma dinâmica evolutiva diferencial além de se acumulam em regiões espécies específicas, sendo verificadas a nível cromossômico pelo FISH, podendo contribuir com a diversificação dos cromossomos entre os organismos.

INTRODUÇÃO

Elementos genéticos móveis (TEs) são seguências de DNA repetitivo com capacidade de se transpor nos genomas e que são encontrados clusterizados ou dispersos nos cromossomos. Três modelos explicam essa distribuição: (i) ruptura de genes, quando ocorre silenciamento gênico; (ii) mecanismo de recombinação ectópica, na qual ocorre recombinação entre cromossomos heterólogos devido ao fato de as sequências serem repetitivas e; (iii) preferências de locais de inserção, que podem controlar a atividade do elemento (Cáceres et al., 2001; McVean 2010). Os TEs podem ser transmitidos pelo mecanismo de transferência horizontal, entre espécies que não são relacionadas transmitidas filogeneticamente ou verticalmente. entre espécies filogeneticamente relacionadas e, a partir disso, passam a seguir caminhos evolutivos independente em cada um desses genomas.

Os TEs podem influenciar na evolução genômica tendo relação com a manutenção, estrutura e funcionamento dos genomas, isso ocorre devido ao fato de o mecanismo de transposição ter capacidade de promover rearranjos cromossômicos, atuar na regulação da expressão gênica ou servirem como substrato evolutivo para processos de domesticação molecular (Feschotte et al, 2007; Hollister e Gaut, 2009; Casacuberta e Gonzalez, 2013), um exemplo disto é a evidência de TEs domesticados molecularmente no genoma do anfíbio *Xenopus tropicalis*. Nesta espécie sugere-se que o elemento retroviral endógeno XTERV1, muitas vezes classificados como retrotransposons com LTRs, é homólogo a proteína FR47 de *Rana sylvatica*, que é regulada pela exposição ao frio, impedindo o congelamento dos organismos (Sinzelle et al., 2011).

O conteúdo de elementos transponíveis (TEs) de um genoma, devido a sua dinâmica evolutiva, pode interferir de maneira significativa no aumento do tamanho dos genomas (Elliott e Gregory, 2015). O tamanho do genoma dos anfíbios pode ter relação ao acúmulo de sequências repetitivas. Em salamandras a variação do tamanho do genoma é atribuída à quantidade de sequências repetitivas (Bozzoni e Beccari, 1978; Mable et al, 2011) do tipo SINEs (Batistoni et al, 1995; Vignali e Nardi, 1996) e aos LINEs pertencentes à família Gypsy/Ty3 na salamandra *Hydromantes* (Marracci el al., 1996). Segundo Sun et al (2012), que analisaram os dados de seis espécies de Plethodontidae, a maior família de salamandra, verificou que os TEs são abundantes e diversificados nesses genomas, sendo os LTRs os mais abundantes, o que contribuiu para os genomas destas salamandras serem gigantes.

Dentre os TEs presentes no genoma de anuros, os retrotransposons da ordem DIRS formam um grupo de elementos muito interessante do ponto de vista evolutivo. Esses elementos codificam uma enzima chamada tirosina recombinase (YR) em vez de uma integrase ou endonuclease e apresentam de modo geral três ORFs (Cappello et al., 1985). As repetições das extremidades desses elementos, podem ser do tipo repetições terminais invertidas (terminal inverted repeats -ITRs) ou repetições diretas (split direct repeats - SDRs) (Cappelo et al., 1985; Goodwin et al. 2004; Poulter and Butler 2015).

Gazolla et al. (2022) identificaram e compararam os elementos de ordem DIRS no genoma de *X. tropicalis e X. laevis*, sendo que a essas análises corroboram com a descrição dos elementos, com relação as repetições, ORFs e domínios para as superfamílias *DIRS-like* e *Ngaro-like*, da ordem DIRS, sendo que esse último foi reportado pela primeira vez para *X. laevis* (Goodwin e Poulter, 2001; Poulter e Goodwin, 2005). Gazolla et al. (2022) identificaram ondas recentes de amplificação dos elementos em ambos os genomas, indicando que isso pode ter sido o fator de diversificação das famílias dos elementos da ordem DIRS nos genomas. Por fim, Gazolla et al. (2022) identificaram evidências de atividade transcricional através dos dados de transcriptoma em *X. tropicalis* e pela primeira vez, em *X. laevis*, através da biblioteca de EST das espécies.

Do ponto de vista cromossômico, apesar do longo tempo de divergência (aproximadamente 136 milhões de anos), P. carvalhoi e X. tropicalis parecem ter retido a condição plesiomórfica da família, ambas com cariótipo 2n = 20 cromossomos (Mezassalma et al., 2015) e morfologia semelhante para alguns pares cromossômicos (Zattera et al., 2019). Já o mapeamento de motivos microssatélites sugere que P. carvalhoi e X. tropicalis compartilham motivos CAG, CGC e GATA nas regiões terminais e subterminais dos cromossomos e também GAA e GATA no par 5 (Zattera et al., 2020), além de marcações específicas para cada uma das espécies, como o acúmulo de GATA no braço longo do par cromossomo 1 de P. carvalhoi e X. tropicalis acúmulo de CA. Pipa carvalhoi apresentou outras marcações específicas, como o GA com maior evidência na região subterminal dos pares 2 e 9, além de marcações CAG no braço longo do par 2 e centromérica do par 6, CCG na região centromérica do par 1 e 4. O par 4 de X. tropicalis apresenta acúmulo de GATA e GA. A presença de sequências teloméricas intersticiais (ITS) nos cromossomos dos pares 6, 8 e 9 de *P. carvalhoi* sugerem possíveis relíquias de inversões pericêntricas responsáveis pela alteração na morfologia desses pares (Zattera et al., 2020) e apontam para a importância dos DNAs repetitivos na diversificação cariotípica entre essas espécies.

No genoma de *X. tropicalis* os TEs correspondem a cerca de 33% de seu conteúdo (Hellsten et al., 2010) tanto de retrotransposons (LTR e não-LTR) quanto de transposons de DNA, sendo esta a classe de maior abundância. O genoma de *X. laevis* foi originado de um processo de alopoliploidia, com tamanho estimado de 3,1 Gb sendo possível designar dois subgenomas, chamados S (shorter) e L (longer) (Matsuda et al., 2015; Session et al., 2016). Curiosamente, estudos genômicos têm revelado TEs específicos para cada subgenoma, por exemplo, XI-TpL_harb e XI-TpS_harb, que são subfamílias de elementos transponíveis MITE (*miniature inverted repeat transposable elements*), exclusivos dos subgenomas L e S, respectivamente. Além disso, elementos TC1 mariner, XI-TpS_mar foram identificados somente no subgenoma S (Session et al., 2016).

Outro MITE denominado de miDNA4 também foi recuperado no genoma de *X. tropicalis*, tendo 79% de similaridade com os de *X. laevis* (Scalvenzi e Pollet, 2014). A superfamília de PiggyBac denominada TxpB também é compartilhada por *X. tropicalis, X. laevis e X. borealis*, onde é possível de se detectar as subfamílias, Kobuta, Uribo1 e Uribo2 (Hikosaka et al., 2007). Alta similaridade, a nível de famílias de TEs, entre espécies de anuros altamente divergentes, também foi verificada, como o caso de *Oophaga pumilio e X.laevis*, recuperando cópias de TEs das famílias Gypsy, Mariner e Tc1, exibindo até 82% de similaridade em algumas famílias (Rogers et al., 2018).

Pipa carvalhoi, o único representante sul americano da família Pipidae, foi recentemente incluída em análises genômicas (Gama et al., 2022). Estima-se que *P. carvalhoi* tenha divergido de *X. tropicalis* e *X. laevis* a pelo menos 136 milhões de anos (Canatella 2015). Analisando especificamente retrotransposons da ordem LINE, Gama et al (2022) identificaram clados/superfamílias de elementos compartilhados com *X. tropicalis* e *X. laevis*, sugerindo sua transmissão vertical a partir do último ancestral comum de Pipidae e retidos no genoma destas espécies. Interessantemente, esses autores encontraram que alguns elementos pertencentes ao grupo Rex1 compartilham grande similaridade de sequências com elementos de alguns peixes, sendo explicadas pelo processo de transferência horizontal (Gama et al., 2022).

Comparar o conteúdo de TEs de espécies filogeneticamente relacionadas permite entender o papel dessas sequências na evolução dos genomas desde sua divergência a partir de um último ancestral em comum. Um exemplo disso são as comparações do conteúdo de TEs entre as espécies *X. tropicalis* e *X. laevis* (Pipidae, Anura), linhagens evolutivas que divergiram há pelo menos 48 milhões de anos (Canatella 2015; Session et al. 2016). Neste trabalho, combinamos buscas genômicas e mapeamento cromossômico de sequências homólogas de retrotransposons das superfamílias *DIRS-like* e *Ngaro-like*, analisadas por Gazolla et al. (2022), em *P. carvalhoi e X. tropicalis* em busca de (i) evidências sobre sua manutenção evolutiva dessas sequências em *P. carvalhoi* apesar da profunda divergência evolutiva entre as espécies e (ii) de sua participação na diversificação destes cariótipos.

MATERIAIS E MÉTODOS

1. Buscas genômicas por sequências homólogas aos elementos das superfamílias DIRS-like e Ngaro-like em P. carvalhoi

Utilizamos como queries as sequências de aminoácidos da transcriptase reversa (RT) dos elementos DIRS-37_XT, DIRS-53_XT e DIRS-54_XT do genoma de *X. tropicalis*, disponíveis no banco de dados Rebpase (Jurka, 2000) versão 23.11 e que compuseram a biblioteca de sequências utilizadas por Gazolla et al. 2022. A escolha por estas famílias foi em virtude de DIRS-37_XT e DIRS-53_XT serem elementos completos (ORFs, domínios conservados e repetições terminais), além de serem representantes, respectivamente, da superfamília *DIRS-like* e *Ngaro-like*. O elemento DIRS-54_XT foi utilizado devido ao fato de ter somente duas superfamílias de elementos do tipo *Ngaro-like* depositado no REPBASE, além disso, esse elemento estava completo em relação as ORFs e domínios conservados, porém não apresentava as repetições terminais. A partir deste momento, quando nos referenciarmos ao elemento DIRS-53_XT ele será tratado como *Ngaro-like*.

Scripts, em linguagem *python*, fornecidos por Gama et al. 2022 foram usados para recuperar as sequências do genoma de *P. carvalhoi* pelo método de Blast Local, realizado offline, que deve seguir algumas etapas: (i) formatação do genoma; (ii) escolha do blast a ser utilizado; (iii) filtragem dos dados; (iv) recuperar as sequências de transposons (Anexo I – comandos utilizados). O genoma de *P. carvalhoi* foi sequenciado pelo Laboratório de Citogenética Evolutiva e Conservação Animais da Universidade Federal do Paraná (UFPR) – Brasil. A plataforma de sequenciamento utilizada por Illumina HiSeq-400 com cobertura de 46x e as leituras foram cortadas com Trimmomatic (Bolger et al., 2014) e montadas com Velvet 1.2.10 (Zerbino e Birney, 2008) pelo método *de novo*, com valor de 105 para o comprimento de hash (kmer). Atualmente o genoma se encontra em nível de contig, com 5.029.551 contigs e 2,4Gb de tamanho. O hardware utilizado na montagem do genoma possui 64 núcleos e 8 processadores Intel ® Xeon ® CPU E5-4640, 2,4 Ghz e 512 de memória, pelo sistema SGI UV100. A montagem do genoma encontra-se disponível em http://200.236.3.93/xeno/velvet/contigs_k105.fa. Neste trabalho, realizamos

uma filtragem dos contigs, utilizando comente os maiores de 400 pb, totalizando 1.153.690 *contigs* e 1,3 Gb de sequências.

Foi utilizado um pipeline, denominado PiRATE (Berthelier et al., 2018), disponível para download em http://www.seanoe.org/data/00406/51795/, desenvolvido para detectar, classificar e anotar TEs presentes em um genoma e verificar quais deles são potencialmente autônomos ou não autônomos. O PiRATE é utilizado em três fases, combinando múltiplas ferramentas e bases de dados disponíveis e realizando a análise dos *contigs* baseado em similaridade, estrutura e repetibilidade das sequências. A etapa de detecção é realizada por meio de quatro ferramentas, das quais utilizamos três, sendo a primeira o Repeat Masker, que detecta com base na similaridade, portanto, necessita de um banco de dados TE, neste caso utilizamos o banco de dados de vertebrados, para fornecer a estimativa de conteúdo; a segunda ferramenta, MITE-hunter, é baseada na estrutura; a terceira ferramenta chamada Repeat Scout é baseada na repetibilidade das cópias.

Na etapa de classificação, é utilizado um banco de dados de TEs, o Repbase, e um programa denominado PASTEC (disponível em: https://urgi.versailles.inra.fr/Tools/PASTEClassifier/PASTEClassifier-tuto) que classifica os TEs com base na classificação de Wicker et al (2007) e agrupa as sequências em potencialmente autônomas, não autônomas e não categorizadas.

A análise molecular estrutural foi baseada em Gazolla et al. (2022), que consiste na identificação dos ORFs pelo programa e NCBI "Open Reading Frame Finder" (ORFfinder) (https://www.ncbi.nlm.nih.gov/orffinder/) com os parâmetros ("ORF mínimo comprimento (nt) "= 75;" Código genético ": 1. Padrão;" códon de início de ORF a ser usado ": somente ATG); os domínios conservados foram analisados pela ferramenta NCBI" Conserved Domains Search Service "(CD-Search) (Marchler-Bauer e Bryant 2004) com um limite de valor e ajustado para 0,1; e a presença das sequências nas extremidades terminais foi verificada a fim de identificar a presença de repetições invertidas terminais (ITRs), região complementar interna (ICRs) e split direct repeats (SDRs) foi investigado usando NCBI BLASTn com a mesma sequência como query e subject, selecionando as opções "Align two or more sequences" (blastn).

2. Extração de DNA de Pipa carvalhoi e Xenopus tropicalis

A extração do DNA genômico se deu pelo método de TNES. O DNA extraído foi quantificado em picodrop e armazenado no freezer.

3 - Isolamento das cópias de DIRS- like e Ngaro-like presentes nos genomas

Utilizamos a reação em cadeia da polimerase (PCR) para amplificar cópias dos elementos DIRS-like e Ngaro-like do genoma de cada espécie. Para isso, utilizamos as sequências de consenso de DIRS-37 XT (representante da superfamília DIRS-like) e DIRS-53 XT (representante da superfamília Ngaro-like) de X. tropicalis depositado no banco de dados Repbase, para a escolha dos primers utilizados. Os primers foram desenhados através do programa Primer-Blast (disponível em: https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Os primers foram concebidos para isolar um fragmento de aproximadamente 351 pb da região da transcriptase reversa das cópias DIRS-like [(F 5 'TACCCTTCGGGCTCTCATCA 3') + (R 5 ۲ de CCTGTAGGGTCTGCATGGTC 3')] e 354 pb da RT de cópias Ngaro-like [(F 5 'TGAGGCAGCTTTCAGGCTAT 3) + (R 5' ACTTGGCATGTTGTAGTGCG 3 ')].

As reações de amplificação foram realizadas usando: 20-40 ng de DNA genômico, 0,2ul de primer F, 0,2ul de primer R, 0,16 mM de dNTPs, 1U de Taq DNA polimerase e 1,5 mM de MgCl2 em 1 × tampão de reação (200 mM de Tris, pH 8,4, 500 mM KCL). O programa de amplificação do tipo DIRS-*like* consistiu nas seguintes etapas: 5 min-95°C / (30s-95°C / 30s-56°C / 2min-72°) x 35 ciclos / 7min-72°C. Já o programa de amplificação do Ngaro-*like* consistiu em: 5 min-94°C / (30s-94°C / 30s- 56°C / 1 mim-72°) x 35 ciclos / 10 min-72°C Ambos os produtos de PCR foram visualizados em eletroforese em gel de agarose 1 % e recuperou bandas únicas com aproximadamente 350 bp para ambos os fragmentos *DIRS-like* e *Ngaro-like*, de acordo com nossas previsões.

4 - Purificação e clonagem das cópias de DIRS- like e Ngaro-like

Os amplicons foram purificados usando o kit EasyPure Quick Gel Extraction (PROMEGA) para remover dNTPs e primers não incorporados. Eles foram inseridos no vetor de clonagem pJET 1.2 / blunt. Realizamos a transformação de células de *Escherichia coli* TOP10 usando a técnica de DNA recombinante. Todos os clones recombinantes foram cultivados e posteriormente fizemos a extração do DNA plasmidial pela técnica de miniprep baseada em Sambrook et al. (1989). Os clones de ambas as espécies foram recuperados para as duas famílias analisadas (DIRS-37_XT e DIRS-53_XT).

5- Hibridização in situ de fluorescência (FISH)

Suspensões celulares

As suspensões celulares de *P. carvalhoi* foram previamente preparadas por Zattera et al., (2019) de espécimes coletados em Buíque, Pernambuco, Brasil (08 ° 37'23 " S, 37 ° 09'21 " W). As preparações de células intestinais e testiculares seguiram o método proposto por King e Rofe (1976) e Schmid (1978). A coleta dos espécimes foi autorizada pelo SISBIO / Instituto Chico Mendes de Conservação da Biodiversidade (protocolo 55481-1), e os espécimes foram depositados no Museu de História Natural da Universidade Federal de Alagoas (MHN-UFAL) em Maceió, Brasil. Já as suspensões celulares de X. tropicalis foram fornecidas pelo Dr. Nicolas Pollet do Laboratoire Evolution Genomes Comportement Ecologie, da Université Paris-Saclay, França. Estas suspensões foram obtidas a parir de culturas de fibroblastos (Sinzelle et al., 2012). Deste cultivo, duas populações celulares foram obtidas, sendo uma com cariótipo 2n=20 (característica de *X. tropicalis*) e a segunda com cariótipo 2n=21, gerada pela presença de uma trissomia do cromossomo 10, sendo esta última linhagem (estável) denominada "Speedy" (Sinzelle et al., 2012).

Para investigar a distribuição cromossômica dos elementos-alvo as sondas produzidas a partir de fragmentos de seguências isoladas dos genomas de X. tropicalis e P. carvalhoi foram marcadas com digoxigenina-11-dUTP (Roche) usando o kit PCR Dig Probe Synthesis (Roche). Utilizamos uma solução contendo 2,5uL de Tampão 10x; 3uL de MgCl2 50mM; 1uL de mix de dNTP 4mM; 0,5uL dUTP-dig 1mM; 1uL de primer foward; 1uL de primer reverse; 2 uL do produto de PCR; 1 uL de água e completamos com água milliQ para ficarmos com uma concentração final de 25uL. Fizemos como controle nucleotídeo marcado. uma reação sem 0 0 programa utilizado no termociclador compreendeu as seguintes etapas: 94°C por 5 minutos; 94°C por 30 segundos; 64°C por 30 segundos; 72°C por 1 minuto; os passos de 2 a 4 foram repetidos por 30 vezes; posteriormente mais um passo a 72°C por 10 minutos, finalizando a 4°C até o armazenamento do produto. Os produtos de amplificação foram observados em gel de agarose a 1%.

O produto de PCR marcado foi precipitado com DNA de espermatozoide de salmão sonicado (100 ng / 1ul) usando acetato de sódio 3M (1/10 de volume) e etanol. O sedimento foi lavado em etanol a 70% e ressuspenso em tampão de hibridização (50%)

formamida, 2x SSC e 10% de sulfato de dextrano). O protocolo de hibridização foi realizado de acordo com Viana et al., 2019. As sondas marcadas com digoxigenina foram detectadas utilizando um anticorpo antidigoxigenina conjugado com rodamina (Roche), seguindo as instruções do fabricante.

6 - Hibridização genômica comparada (CGH) em P. carvalhoi e Xenopus tropicalis

Comparações interespecíficas do conteúdo repetitivo pelo método CGH, foram aplicadas para confirmar o enriquecimento cromossômico espécie-específico de TEs. O genoma total (gDNA), derivado das duas espécies, foi hibridizado com os cromossomos metafásicos de ambas. Essa abordagem nos permite comparar os gDNAs de duas espécies (ou mais) em busca de distribuição diferencial de classes repetitivas específicas nos genomas divergentes e sua distribuição cromossômica (Chester et al. 2010).

O gDNA de *X. tropicalis* foi marcado com biotina-16-dUTP e o gDNA de *P. carvalhoi* com digoxigenina-11-dUTP por meio da técnica de Nick *translation* (Roche, Mannheim, Alemanha). Para a comparação interespecífica, a mistura final de hibridização para cada lâmina foi composta pelo gDNA de cada espécie (500ng cada), 20ug de DNA C0t-1 das espécies analisadas (DNA genômico com sequências altamente repetitivas preparadas de acordo com Zwick et al. (1997)) e 20ul do tampão de hibridização contendo 50% de formamida, 2 × SSC, 10% de SDS, 10% de sulfato de dextrana e solução de Denhardt, pH 7,0. O FISH seguiu o protocolo baseado em Symonová et al. (2015) com algumas modificações. As lâminas foram incubadas a 37°C em câmara úmida e escura por três dias. Para detectar os sinais, foram usados antidigoxigenina-rodamina (Roche) e avidina-FITC (isotiocianato de fluoresceína, Sigma), o primeiro foi diluído em albumina sérica bovina (BSA) a 0,5% em solução salina tamponada com fosfato (PBS) e o segundo PBS contendo 10% de soro de cabra normal (NGS). Os cromossomos foram corados com DAPI (1,2 μg / ml) e montados em solução antifade (Vector, Burlingame, CA, EUA).

RESULTADOS

1. Caracterização molecular dos elementos da ordem DIRS no genoma de Pipa carvalhoi

O conjunto de *contigs* do genoma de *P. carvalhoi* foi submetido ao pipeline PiRATE, sendo recuperados 944.532 fragmentos com alguma assinatura de TEs. A partir disso, uma nova filtragem foi realizada para remover novamente os fragmentos menores que

400 bp resultando em um total de 67.847 fragmentos. Destes, a ferramenta PASTEC foi capaz de identificar 50.472 fragmentos de retrotransposons, 10.557 de transposons de DNA, 6 genes potenciais do hospedeiro, 4.145 SSR e 2.667 não categorizados (Tabela 1).

Tabela 1: Conteúdo dos elementos de transposição encontrados apartir do conjunto de contigs de tamanho >400 pb no genoma de *Pipa carvalhoi*

Sequência Repetitiva	Quantidade		
Retrotransposons	50.472		
Transposons de DNA	10.557		
Genes potenciais	6		
SSR	4.415		
Não categorizados	2.667		
Total	67.847 fragmentos		

Foram recuperados um grupo de nove cópias de elementos *DIRS-like* e vinte cópias para *Ngaro-like* quando aplicamos os mesmos critérios de filtragem de cópias de acordo com Gazolla et al., (2022). Dez cópias foram recuperadas usando as sequências de aminoácidos da RT de DIRS-37_XT (*DIRS-like*) como *query* e 20 cópias usando as sequências de NGARO-53_XT e NGARO-54_XT (*NGARO-like*) (Tabela 1 – Material Suplementar). -

As sequências recuperadas da superfamília *DIRS-like* variaram em tamanho de 420 pb (DIRS-37_XTcopy5PC) a 6144 pb (DIRS-37_XTcopy1PC). Analisamos as três ORFs em todas as famílias e identificamos a presença da ORF1, que apresenta a sequência da Gag-*like*, com o domínio LAP2alpha, recuperado nas cópias 2, 3 e 6 (Tabela 1 - Material Suplementar). Quanto à ORF 2, que possui os domínios RT e RH, apenas foram identificados nas cópias 1,2,4,6,7,8 e 9 (Tabela 1 – Material Suplementar). Nenhuma das cópias analisadas apresentou o domínio desoxi-adenosina metilase (DAM) que pode aparecer na ORF 2 desses elementos. Em nenhuma das cópias analisadas identificamos a ORF 3, correspondente a proteína tirosina recombinase (YR). Todos os *contigs* estão incompletos e não mostraram as repetições terminais. Portanto, as ORFs 1 e 2, foram as que apresentaram o maior número de regiões conservadas entre os elementos da superfamília *DIRS-like* (Tabela 1 – Material Suplementar), isso pode ter

devido a montagem do genoma ou pela estratégia de busca que realizamos, pois utilizamos como *query* a sequência da RT que está localizada na ORF2 dos elementos.

Na família *Ngaro-like*, as cópias recuperadas também foram analisadas seguindo a estrutura molecular previamente identificada em *X. tropicalis* e *X. laevis*, onde as repetições são do tipo diretas (SDRs) (Gazolla et al, 2022). São quatro repetições que caracterizam esses elementos, sendo denominadas A1 e A2 e B1 e B2. As ORFs seguem o padrão semelhante a *DIRS-like*, mas apresenta uma ORF adicional, posterior a ORF 3, que apresenta um domínio que codifica para a proteína SGNH_hidrolase. Recuperamos quatorze sequências semelhantes a *Ngaro-like* do genoma de *P. carvalhoi*. O tamanho das sequências recuperadas variou de 424 bp a 6736 bp. Todos os *contigs* estão incompletos e nenhuma repetição terminal foi observada em nenhuma sequência. A ORF1 está presente nas cópias 2, 4, 5 e 8 recuperadas a partir de *Ngaro*-53_XT. Apenas os RTs das cópias 2, 3 e 4 foram recuperados nas cópias de *Ngaro*-53_XT. Já dentre as cópias recuperadas a partir de *Ngaro*-54_XT, as dez sequências apresentavam a ORF 2 e a sequência do domínio conservado das RTs de maneira completa, porém somente a cópia 9 mostrou o domínio conservado da RH (Tabela 1 – Material Suplementar).

2. FISH com sonda DIRS-37_XT e NGARO-53_XT em P. carvalhoi e X. tropicalis

Em *X. tropicalis*, com a sonda DIRS-37_XT, sempre analisando pelo menos dez metáfases, detectamos sinal de hibridação em praticamente todo o braço curto dos cromossomos do par 3. Observamos marcações regiões subterminais dos homólogos dos pares 1, 2, 6, 8 e um cromossomo do par 10, em ambos os braços cromossômicos (Figura 1). Com a sonda *Ngaro*-53_XT detectamos sinais de hibridação na região intersticial do braço longo dos homólogos do par 7 (Figura 2). Na região subterminal, os pares 1, 2, 3, 5, 8 e 10 exibem sinais mais evidentes, porém os pares 1 e 2 os sinais aparecem em apenas um dos homólogos, enquanto que nos pares 3, 5, 8 e 10, o sinal é evidente em todos os homólogos (Figura 2).

Figura 1: Cariograma correspondente a *X. tropicalis* submetido à FISH com sonda DIRS-37_XT (*DIRS-like*) marcada com digoxigenina-11-dUTP, na suspensão do tipo *Speedy*, contendo a trissomia do cromossomo 10. O padrão de marcação sempre foi observado em pelo menos dez metáfases para cada sonda.



Figura 2: Cariograma correspondente a *X. tropicalis* submetido à FISH com sonda NGARO-53_XT (*Ngaro-like*) marcada com digoxigenina-11-dUTP, na suspensão do tipo *Speedy*, contendo a trissomia do cromossomo 10. As hibridizações correspondem às marcações na coloração rosa presente nos cromossomos.



Os ensaios de FISH utilizando as sondas DIRS-37_XT e NGARO-53_XT amplificadas de *X. tropicalis*, não apresentaram sinais de hibridação em nenhum dos cromossomos do cariótipo de *P. carvalhoi* (Figura 3 e Figura 4, respectivamente). Os experimentos foram refeitos com sondas amplificados a partir do genoma de *P. carvalhoi* e também não obtivemos resultados positivos. Para cada sonda, os experimentos foram realizados sempre utilizando como controle positivo do experimento a sonda DIRS-37_XT no cariótipo de *X. tropicalis*, totalizando oito testes e nenhum deles apresentou marcação.

Figura 3: Metáfase de *Pipa carvalhoi* submetida à FISH com sonda DIRS-37_XT (*DIRS-like*) marcada com digoxigenina-11-dUTP. (A) cromossomos corados com DAPI, (B) detecção da sonda DIRS-37_XT marcada com digoxigenina-11-dUTP e (C) sobreposição das capturas de imagens. Note que não há sinal de marcações da sonda nos cromossomos.



Figura 4: Metáfase de Pipa carvalhoi submetida à FISH com sonda NGARO-53_XT (NGARO-like) marcada com digoxigenina-11-dUTP. (A) cromossomos corados com DAPI, (B) detecção da sonda NGARO-53_XT marcada com digoxigenina-11-dUTP e (C) sobreposição das capturas de imagens. Note que não há sinal de marcações da sonda nos cromossomos.



3. CGH em Pipa carvalhoi e Xenopus tropicalis

Utilizando a metodologia de CGH, buscamos validar se os pontos únicos de acúmulo de sequências *DIRS-like* e *Ngaro-like* em *X. tropicalis* realmente representavam blocos cromossômicos espécies-específicos deste genoma, especialmente os sinais de acúmulo já mencionados da sonda DIRS-37_XT no braço curto do par 3 (Figura 1) e Ngaro-53_XT na constrição secundária do braço longo do par 7 (Figura 2).

Utilizando o genoma de P. carvalhoi marcado com biotina e de X. tropicalis marcado com digoxigenina-11-dUTP, hibridizado em cromossomos metafásicos de X. tropicalis, nós detectamos que as regiões dos pares de cromossomos três e dez têm uma envolvendo todo o marcação específica braço curto desses cromossomos subtelocêntricos. Além disso, uma marca muito evidente aparece na região intersticial do braço longo do par sete. As marcações nos pares 3 e 7 são coincidentes com as verificadas para a sonda DIRS-37 XT e NGARO-53 XT, respectivamente. Regiões que compartilham o genoma de P. carvalhoi aparecem nos outros pares de cromossomos, principalmente nas regiões subterminais mostrando que há homologia de sequências repetitivas nessas regiões, visto que são regiões enriquecidas de DNAs repetitivos que possivelmente sejam compartilhadas entre essas duas espécies (Figura 5).

Em *P. carvalhoi*, a região centromérica dos pares de cromossomos um, três e dez se mostra espécie-específica (Figura 6). Os pares de cromossomos dois, quatro, cinco, seis, sete, oito e nove, tanto nas regiões pericentromérica quanto subterminal, mostram que há uma sequência compartilhada com o genoma de *X. tropicalis*, pelo fato do sinal compartilhar a marcação gDNA de *P. carvalhoi* e de *X. tropicalis*.
Figura 5: Cromossomos metafásicos de *X. tropicalis* submetidos a experimentos de CGH. O genoma de *P. carvalhoi* foi marcado com digoxigenina-11-dUTP (vermelho), enquanto o genoma de *X. tropicalis* foi marcado com biotina (verde). As regiões sobrepostas são exibidas em amarelo em cromossomos metafásicos de *X. tropicalis*.



Figura 6: Cromossomos metafásicos de *P. carvalhoi* submetidos a experimentos de CGH. O genoma de *P. carvalhoi* foi marcado com digoxigenina-11-dUTP (vermelho), enquanto o genoma de *X. tropicalis* foi marcado com biotina (verde). As regiões sobrepostas são exibidas em amarelo nos cromossomos de *P. carvalhoi*.



DISCUSSÃO

Nossas buscas genômicas por superfamílias de elementos *DIRS-like* e *Ngaro-like* em *P. carvalhoi* revelaram a presença de elementos homólogos aos recuperados no genoma de *X. tropicalis* e *X. laevis* (Gazolla et al., 2022), sugerindo que tenham sido transmitidos verticalmente a partir do último ancestral comum compartilhado destas espécies, a pelo menos 136 milhões de anos (Session et al., 2016). Nós identificamos mais de 50 mil fragmentos dos aproximadamente 67 mil recuperados no genoma de *P. carvalhoi* com similaridade à retrotransposons.

Dentre os retrotransposons recuperados a partir do genoma de *P. carvalhoi*, a análise das cópias recuperadas de elementos das superfamílias *DIRS-like* e *Ngaro*-like, não recuperou nenhum dos elementos contendo as repetições terminais características dos elementos da ordem DIRS. Atribuímos este resultado ao fato do genoma de *P. carvalhoi* encontrar-se muito fragmentado, apesar de uma cobertura de 46x, devido à dificuldade de montagem do genoma, tornando difícil a busca por elementos completos. Gama et al. (2022) atribui o fato de uma alta proporção de cópias de TEs recuperadas estarem incompletas à fragmentação do genoma.

Apesar das restrições de análise impostas pelo nível de montagem de nosso genoma, a presença de cópias incompletas também pode ser explicada pelo longo tempo de persistência dessas cópias no genoma a partir de sua inserção, podendo ter resultar em alterações a nível genômico e também cromossômico das espécies. Neste sentido, podemos atribuír esse resultado ao "ciclo de vida" dos TEs nos organismos durante a evolução, visto que, o mesmo, compreende basicamente 3 fases: (i) invasão e amplificação dos elementos, acompanhada de mutações que podem tornar alguns elementos inativos; (ii) uma fase de maturação, em que ocorre um equilíbrio do TE no genoma com relação as amplificações ou perda das cópias; (iii) e a fase de senescência dos TEs pelos mecanismos de silenciamento através da degeneração da sequência dos elementos que causa interrupção de sua potencialidade funcional (Kidwell e Lisch, 2001; Fernandéz-Medina et al., 2012).

Não conseguimos identificar uma assinatura terminal das sequências, portanto, não sabemos se há alguma marcação trinucleotídica diferencial para *P. carvalhoi,* como identificado para *X. tropicalis e X. laevis* a trinca nucleotídica terminal "TTT" (Gazolla et al., 2022). Apesar disso, não afirmamos que as sequências do genoma de *P. carvalhoi* são todas degeneradas, pois não descartamos a hipótese de que o genoma apresenta elementos potencialmente ativos da ordem DIRS, visto que, os vários fragmentos que

conseguimos recuperar, mostram a existência de ORFs potencialmente funcionais, bem como a presença de domínios, dentre eles a RT. O fato de o genoma estar fragmentado, pode ser a explicação para não conseguimos recuperar as sequências completas, incluindo as repetições terminais. Como usamos *scripts* que seleciona a sequência com 3kb *upstream e downstream*, os resultados não nos retornaram elementos completos. Concluímos que nossos achados são consistentes em considerar os elementos descritos para a ordem DIRS, pois demonstra a presença das estruturas moleculares como ocorre em outros organismos (Goodwin e Poulter, 2001; Poulter e Goodwin, 2005; Gazolla et al, 2022), um exemplo disso, é a cópia seis do elemento DIRS-53_XT que recuperamos em *P. carvalhoi*, que apresenta a RT, RH e a YR.

Gazolla et al. (2022) identificou eventos de amplificação dos elementos DIRS-37_XT e NGARO-53_XT, através da análise de landscape, que é um modelo de evolução da paisagem de como os TEs se comportam durante a evolução dos genomas, essa análise é baseada na sequência consenso de cada superfamília e calcula-se a divergência Kimura-2-parâmetros, com CpG ajustado. A amplificação dos elementos pode ter acontecido no genoma de *P. carvalhoi,* onde em determinada época, os elementos da ordem DIRS passaram por um processo de amplificação no genoma e posteriormente degeneração, sem que houvesse uma nova onda recente de amplificação. Isso explicaria o fato de os elementos serem degenerados e não ser apenas um artefato da montagem do genoma ou da estratégia de busca. O fato de não termos identificado elementos completos, não indica que estão todos degenerados, visto que, essa análise de landscape necessita de uma sequência consenso baseada na biblioteca de elementos da espécie, que neste caso, não pode ser executada até o momento.

O padrão de acúmulo de DNA repetitivo dos sinais detectados pela técnica de FISH entre os genomas, foi observada na maioria das extremidades dos cromossomos metafásicos de *X. tropicalis,* como visto também pela grande quantidade de elementos repetitivos nessas regiões, principalmente os motivos microssatélites detectados pelas sondas GA, CA, GATA, CAG e CCG (Zattera et al., 2020). Esse padrão de hibridização é congruente com as análises de marcadores e utilizando FISH com sequências de microssatélites e COT1 (*fraction of repetitive DNA sequences*) que é uma fração de DNA repetitivo do DNA competidor junto ao DNA genômico algo, para impedir que ocorra hibridação não específica (Zattera et al., 2020; Roco et al., 2021). A ocorrência de marcações de microssatélites espécies-específicas encontradas por Zattera et al. (2020) em *P. carvalhoi e X. tropicalis*, reforça a hipótese de evolução independente do conteúdo de DNA repetitivo durante a evolução dos cromossomos (Hartley e O'Neill, 2019).

Portanto, nossos dados concordam a hipótese de Zattera et al. (2020) de que existe a um padrão de distribuição das sequências repetitivas distinto entre as espécies *P. carvalhoi* e *X. tropicalis.*

Apesar dos elementos das superfamílias DIRS-like e Ngaro-like estarem presentes no genoma de *P. carvalhoi*, isso foi verificado pelos contigs recuperados, nossas evidências de mapeamento cromossômico não se deram positiva, porém, não podemos afirmar que as cópias tenham sido perdidas ou degeneradas após a separação destas espécies. Já era esperado que a sonda heteróloga tivesse um resultado negativo para a marcação do FISH, porém a sonda homóloga também não mostrou marcações positivas. Alternativamente, uma explicação para os resultados obtidos nos experimentos de FISH seria o fato de que o limite de detecção pela técnica é de 1.5kb, desta forma realizamos repetições do experimento para nos certificarmos de que realmente a quantidade das cópias dos elementos DIRS-37 XT e DIRS-53 XT são imperceptíveis para P. carvalhoi. Outra questão, é o fato de que os primers utilizados para amplificação foram baseados no genoma de X. tropicalis, e assim, o primer ser heterólogo resultaria em algumas situações distintas: (i) não amplificar no genoma de P. carvalhoi, o que não foi o caso em nosso estudo; (ii) amplificar o elemento, mesmo o primer sendo heterólogo, neste caso resultaria em uma sonda homóloga; (iii) amplificar ao acaso alguma região distinta, sendo necessário o sequenciamento para verificarmos qual tipo de sequência foi amplificada. . Por fim, pode ser que o genoma de *P. carvalhoi* tenha cópias intactas dos elementos, porém se eles estiverem dispersos nos cromossomos, possivelmente o resultado obtido nos experimentos de FISH seriam os mesmos.

O cariótipo de *X. tropicalis,* apresenta intensos sinais, de *DIRS-like,* no braço curto do par cromossômico 3 e 10 e nos leva a acreditar que são possíveis regiões marcadoras moleculares da espécie e reforçando nossa hipótese. Roco et al. (2021) observou no braço curto do par cromossômico 3, uma marcação CMA3 positiva, o mesmo foi observado por Tymowska et al. (1982) em *X. tropicalis.* Zattera et al. (2020) identificou fortes marcações utilizando a sonda (CCG)10 na região do braço curto do par 3 e de CA e GA nas regiões subterminais do par 10 de *X. tropicalis* mostrando que essa região é uma região enriquecida de DNAs repetitivos e uma possível região espécie-específica. Roco et al. (2021) identificou sinais positivos, utilizando como sonda o gDNA de *X. laevis,* nos braços curtos dos cromossomos 3L e 3S da espécie, porém, segundo os autores, as regiões do cromossomo 3 de *X. tropicalis* não tem relação com a marcação dos cromossomos 3L e 3S da espécie, indicando uma evolução independente desses genomas.

O acúmulo de cópias em regiões cromossômicas específicas no cariótipo de *X. tropicalis* das superfamílias *DIRS-like* e *NGARO-like* parecem estar de acordo com os resultados de landscape obtidos por Gazolla et al (2022), revelando a grande concentração de cópias de elementos com alto grau de similaridade genética entre si. Este dado é congruente com os resultados obtidos em nossos experimentos de FISH, considerando o grau de alta estringência utilizado (70%). Por exemplo, o acúmulo de sinais de hibridação na região dos braços curtos dos homólogos do par 3 em *X. tropicalis* são coincidentes com marcações GC-positiva (Tymowska et al. 1982; Roco et al., 2021) e contribuem para entender a composição de sequências dessas regiões.

Outro achado que nos chamou atenção foi a presença de acúmulo de sinal e hibridação da sonda NGARO-53 XT na região da constrição secundária do braço longo dos homólogos do par 7 em X. tropicalis. O cromossomo 7 de X. tropicalis tem sido alvo de investigações citogenômicas por conter o locus de determinação sexual nesta espécie (Seifertova et al 2013; Roco et al., 2015), localizado na região 7p. O acúmulo de Ngaro*like* na região de constrição secundária dos braços longos dos homólogos do par 7 parece ser coincidente com a NOR, detectada tanto por Ag-NOR quanto por FISH (Roco et al., 2021). A utilização do método de CGH revela a região 7q como sendo espécie-específica, composta por um arranjo complexo de sequências repetitivas de natureza distinta que Roco et al. (2021) denominou de XTR-7W para a sonda de todo o cromossomo 7 e XTR-7p para a sonda correspondente ao braço curto do cromossomo 7. Nossos dados de mapeamento cromossômicos correlacionados às análises citogenéticas de outros autores (Roco et al., 2020) não nos permitem propor nenhuma hipótese sobre como estes repetitivos estão arranjados nesta região e futuros estudos incluindo o screening destas porções genômicas por bioinformática e experimentos de FISH em fibras cromatínicas (fiber-FISH) poderão fornecer maior detalhamento dessa região cromossômica.

Variações no sucesso evolutivo de famílias de TEs transmitidas verticalmente entre essas linhagens já foram observadas, como por exemplo, Gama et al. (2022) que recuperou clados de elementos LINEs compartilhados nos genomas de *P. carvalhoi, X. tropicalis e X. laevis*. Existem alguns clados, como o RTE que é compartilhado entre *P. carvalhoi e X. laevis* e o fato de não aparecer em *X. tropicalis*, pode estar relacionado a uma perda completa desta superfamília neste genoma. Também foi recuperado, neste trabalho, outros clados que são espécie-específicos, como por exemplo, L2 e REX1 em *P. carvalhoi.* Mesmo em separações mais recentes, como ocorre com *X. tropicalis e X. laevis*, há diferenças na composição genômica dos TEs, um exemplo é o que foi verificado com os elementos da ordem DIRS por Gazolla et al. (2022) em que as ondas

de amplificação dos TEs se deram em momentos distintos da evolução para espécie e também as superfamílias recuperadas possuem diferenças a nível de sequência conforme a espécie. Esses dados corroboram com a hipótese de que existe uma evolução distinta dos elementos em cada genoma.

Nossos experimentos de CGH também foram capazes de discriminar regiões cromossomo-específicas no cariótipo de *P. carvalhoi* nos cromossomos dos pares 1, 3 e 10. Como utilizamos para a sonda o DNA da própria espécie, o CoT1 se fez necessário como um bloqueio do volume extra de sequências repetitivas que são comuns entre os genomas comparados, neste caso, o genoma de *P. carvalhoi com X. tropicalis,* tendo um alto índice de fidelidade (Peterson et al., 2002) principalmente entre estudos comparativos e evolutivos (Roco et al., 2021), ou seja, a vantagem do CoT1 é que ele aumenta a especificidade de hibridação pois ele é basicamente um *pool* de sequências repetitivas em comum entre os genomas. Desta forma, isso reforça nossa hipótese de que essas regiões são espécie-específicas, pois a grande porção desses segmentos exclusivos para o genoma de *P. carvalhoi* são de microssatélites GA, CA, GATA, CCG e CAG, como já demonstrado por Zattera et al. (2020) que verificou um grande acúmulo dessas sequências repetitivas nos cromossomos desta espécie.

Essas regiões apresentam elementos que no atual momento temporal está específico para determinada espécie. Porém, vale ressaltar que o fato de não termos verificado marcações dos elementos da ordem DIRS em *P. carvalhoi*, não significa que eles não estão presente neste genoma, mas sim que eles se apresentam em menor número incapaz de serem detectados pelo FISH. É de grande interesse, diante da compreensão da presença dos elementos da ordem DIRS no genoma de *P. carvalhoi,* que a análise desses elementos seja ampliada a fim de se compreender o comportamento intragenômico dos TEs nessa espécie, como verificamos em *X. tropicalis* o acúmulo do elemento *DIRS-like* no cromossomo 3 e do *Ngaro-like* na constrição secundária do cromossomo 7.

Finalmente, nossos dados dão suporte a hipótese de que o longo tempo de divergência entre as espécies têm resultado na de divergência cromossômica a partir de acúmulo de DNAs repetitivos, sendo possível identificar segmentos espécie-específicos que se acumularam durante o processo de diversificação devido a evolução independente dos cromossomos. Cenário semelhante já foi reportado em serpentes da espécie *Boa constrictor* na qual foi associado a presença de TEs e SSRs (simple short repeats) influenciando na arquitetura e manutenção do genoma (Viana et al., 2020). Por fim, nosso estudo demonstra que há presença dos TEs no genoma de *P. carvalhoi*, em particular dos

elementos de ordem DIRS pela análise genômica dos fragmentos recuperados. No entanto, o real impacto que os TEs tiveram na diferenciação cromossômica das espécies ainda não foi investigado, para isso será necessário maiores estudos acerca do genoma dessas espécies combinados com técnicas de citogenéticas.

CONSIDERAÇÕES FINAIS

Os resultados de nossos trabalhos reforçam a hipótese de que os elementos de transposição atuam de forma ativa na arquitetura e manutenção dos genomas dos organismos, pois além de estarem presentes nos genomas das espécies que estudamos (X. tropicalis, X. laevis e P. carvalhoi), muitos deles se mostraram potencialmente ativos. As análises da estrutura dos elementos e da paisagem genômica evolutiva, no genoma de X. tropicalis e X. laevis, mostraram que esses elementos se comportam de maneira distinta entre as espécies evoluindo de maneira independente. Além disso, no estudo comparativo entre X. tropicalis e P. carvalhoi, os experimentos de FISH combinados com o de CGH, contribuíram para reforçarmos a hipótese de que as porções repetitivas dos genomas possuem também TEs atuando em regiões espécies-específicas para os organismos, mostrando uma evolução cariotípica distinta, visto que, os elementos se comportam independentemente. Nossos achados reforçam a hipótese de que os elementos de transposição necessitam de estudos que visem não só a sua presença ou ausência nos genomas, mas sim, de análises estruturais, combinadas a citogenética a fim de compreender como eles se comportam nos cromossomos e como influenciam a diversificação das espécies durante a evolução.

REFERÊNCIAS

Batistoni, R.; Pesole, G.; Marracci, S.; Nardi, I (1995). A tandemly repeated DNA family originated from SINE-related elements in the European plethodontid salamanders (Amphibia, Urodela). 40(6), 608–615. doi:10.1007/bf00160508

Beck, C.R.; Garcia-Perez, J.L.; Badge, R.M.; Moran, J.V. (2011) **LINE-1 elements in structural variation and disease.** Annu Rev Genomics Hum Genet 12(1):187–215. https://doi.org/10.1146/annurev-genom-082509-141802

Biémont, C.; Vieira, C. (2006) **Genetics: junk DNA as an evolutionary force.** Nature. 443(7111):521–4.

Bourgeois, Y.; Boissinot, S. (2019) On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. Genes.10(6):419.

Bozzoni, I. ; Beccari, E (1978). Clustered and interspersed repetitive DNA sequences in four amphibian species with different genome size. *520(2)*, *0*– *252*. doi:10.1016/0005-2787(78)90224-1

Brown, J.D.; O'Neill, R.J. (2010) Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. Annu Rev Genomics Hum Genet 11:291–316. https://doi.org/10.1146 /annurev-genom-082509-141554

Cáceres, M.; Puig, M.; Ruiz, A (2001). **Molecular characterization of two natural hotspots in the Drosophila buzzatii genome induced by transposon insertions.** Genome Res 11:1353–1364.

Cannatella, D. (2015). *Xenopus* in space and time: fossils, node calibrations, tipdating, and paleobiogeography. Cytogenet. Genome *Res.* 145, 283–301. doi: 10.1159/000438910

Cappello, J.; Handelsman, K.; Cohen, S.M.; Lodish, H.F. (1985). Structure and Regulated Transcription of DIRS-1: An Apparent Retrotransposon of *Dictyostelium*

discoideum. Cold Spring Harbor Symposia on Quantitative Biology, 50(0), 759–767. doi:10.1101/sqb.1985.050.01.094

Carbone, L.; Alan Harris, R.; Gnerre, S.; Veeramah, K.R.; LorenteGaldos, B.; Huddleston, J.; Meyer, T.J.; Herrero, J.; Roos, C.; Aken, B.; Anaclerio, F.; Archidiacono, N.; Baker, C.; Barrell, D.; Batzer, M.A.; Beal, K.; Blancher, A.; Bohrson, C.L.; Brameier, M.; Campbell, M.S.; Capozzi, O.; Casola, C.; Chiatante, G.; Cree, A.; Damert, A.; de Jong, P.J.; Dumas, L.; Fernandez-Callejo, M.; Flicek, P.; Fuchs, N.V.; Gut, I.; Gut, M.; Hahn, M.W.; Hernandez-Rodriguez, J.; Hillier, L.D.W.; Hubley, R.; Ianc, B.; Izsvák, Z.; Jablonski, N.G.; Johnstone, L.M.; Karimpour-Fard, A.; Konkel, M.K.; Kostka, D.; Lazar, N.H.; Lee, S.L.; Lewis, L.R.; Liu, Y.; Locke, D.P.; Mallick, S.; Mendez, F.L.; Muffato, M.; Nazareth, L.V.; Nevonen, K.A.; O'Bleness, M.; Ochis, C.; Odom, D.T.; Pollard, K.S.; Quilez, J.; Reich, D.; Rocchi, M.; Schumann, G.G.; Searle, S.; Sikela, J.M.; Skollar, G.; Smit, A.; Sonmez, K.; Hallers, B.; Terhune, E.; Thomas, G.W.C.; Ullmer, B.; Ventura, M.; Walker, J.A.; Wall, J.D.; Walter, L.; Ward, M.C.; Wheelan, S.J.; Whelan, C.W.; White, S.; Wilhelm, L.J.; Woerner, A.E.; Yandell, M.;Zhu, B.; Hammer, M.F.; Marques-Bonet, T.; Eichler, E.E.; Fulton, L.; Fronick, C.; Muzny, D.M.; Warren, W.C.; Worley, K.C.; Rogers, J.; Wilson, R.K.; Gibbs, R.A. (2014) Gibbon genome and the fast karyotype evolution of small apes. Nature 513(7517):195-201. https://doi.org/10.1038/nature13679

Casacuberta, E.; González, J (2013). **The impact of transposable elements in environmental adaptation.** Molecular Ecology, 22(6), 1503– 1517. doi:10.1111/mec.12170

Charlesworth, D.; Charlesworth, B.; Marais, G.A.B. (2005) **Steps in the evolution of heteromorphic sex chromosomes.** Heredity. 95, 118–128.

Cioffi, M.B.; Kejnovsky, E.; Bertollo, L.A.C. (2011) The chromosomal distribution of microsatellite repeats in the genome of the wolf fish Hoplias malabaricus, focusing on the sex chromosomes. Cytogenet. Genome Res. 132, 289–296.

Elliot, T.A; Gregory, T.R (2015). Do larger genomes contain more diverse transposable elements? BMC Evol Biol 15:69

Fernández-Medina, Rita D; Ribeiro, José M C; Carareto, Claudia M A; Velasque, Luciane; Struchiner, Cláudio J (2012). Losing identity: structural diversity of transposable elements belonging to different classes in the genome of *Anopheles gambiae*. BMC Genomics, 13(1), 272–. doi:10.1186/1471-2164-13-272

Feschotte, C.; Pritham, E. J. (2007). DNA Transposons and the Evolution of EukaryoticGenomes.AnnualReviewofGenetics,41(1),331–368. doi:10.1146/annurev.genet.40.110405.090448

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. Elsevier Science. 103–107. doi:10.1016/0168-9525(89)90039-5

Frost, D. (2022). **Amphibian Species od the World.** Disponível em: https://amphibiansoftheworld.amnh.org/. Acesso em: Fevereiro de 2022.

Gama, J. M; Ludwig, A.; Gazolla, C. B.; Guizelini, D.; Recco-Pimentel, S. M.; Bruschi, D. P (2022) **A genomic survey of LINE elements in Pipidae aquatic frogs shed light on Rex-elements evolution in these genomes.** Molecular Phylogenetics and Evolution. Volume 168, March 2022, 107393

Gazolla, C. B; Ludwig, A.; Gama, J. M.; Bruschi, D. Evolutionary dynamics of *DIRSlike* and *Ngaro-like* retrotransposons in Xenopus laevis and Xenopus *tropicalis* genomes. G3 Genes|Genomes|Genetics, Volume 12, Issue 2, February 2022, jkab391, <u>https://doi.org/10.1093/g3journal/jkab391</u>.

Gilbert, C.; Feschotte, C. (2018) Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr Opin Genet Dev. 49:15–24

Goodwin, T. J. D.; e Poulter, R. T. M. (2004). **A New Group of Tyrosine Recombinase-Encoding Retrotransposons.** Molecular Biology and Evolution, 21(4), 746–759. doi:10.1093/molbev/msh072

Goodwin, T. J. D.; Poulter, R. T. M. (2001). The DIRS1 Group of Retrotransposons.MolecularBiologyandEvolution,18(11),2067–2082.doi:10.1093/oxfordjournals.molbev.a003748

Hancks, D.C.; Kazazian, H.H. Jr. (2012) Active human retrotransposons: variation and disease. Curr Opin Genet Dev 22:191–203. <u>https://doi.org/10.1016/j.gde.2012.02.006</u>

Hancks, D.C.; Kazazian, H.H. Jr. (2016) **Roles for retrotransposon insertions in human disease.** Mob DNA 7:9. https://doi. org/10.1186/s13100-016-0065-9

Hartley, G.; O'Neill, R (2019). Centromere Repeats: Hidden Gems of the Genome. Genes, 10(3), 223–. doi:10.3390/genes10030223

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

Hikosaka, A., Kobayashi, T., Saito, Y., & Kawahara, A. (2007). Evolution of the Xenopus piggyBac Transposon Family TxpB: Domesticated and Untamed Strategies of Transposon Subfamilies. Molecular Biology and Evolution, 24(12), 2648–2656. doi:10.1093/molbev/msm191

Hollister, J. D.; Gaut, B. S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Research, 19(8), 1419–1428. doi:10.1101/gr.091678.109

Irina, A.; Matthew, M. (2005). Deleterious transposable elements and the extinction of asexuals. 27(1), 76–85. doi:10.1002/bies.20159

Kang, M.I.; Rhyu, M.G.; Kim, Y.H.; Jung, Y.C.; Hong, S.J.; Cho, C.S.; Kim, H.S. (2006) **The length of CpG islands is associated with the distribution of Alu and L1 retroelements.** Genomics 87:580–590. <u>https://doi.org/10.1016/j.ygeno.2006.01.002</u>

Kapitonov, V. V.; Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. Nature Reviews Genetics. 9(5), 411–412. doi:10.1038/nrg2165-c1

Kazazian, H.H. Jr.; Moran, J.V. (2017) **Mobile DNA in health and disease.** N Engl J Med 377:361–370. <u>https://doi.org/10.1056 /NEJMra1510092</u>

Kent, T.V.; Uzunović, J.; Wright, S.I. (2017) **Coevolution between transposable elements and recombination.** Phil. Trans. R. Soc. B 372, 20160458. doi:10.1098/ rstb.2016.0458.

Khokha, M.K.; Krylov, V.; Reilly, M.J.; Gall, J.G.; Bhattacharya, D.; Cheung, C.Y.; Kaufman, S.; Lam, D.K.; Macha, J.; Ngo, C.; Prakash, N.; Schmidt, P.; Tlapakova, T.; Trivedi, T.; Tumova, L.; Abu-Daya, A.; Geach, T.; Vendrell, E.; Ironfield, H.; Sinzelle, L.; Sater, A.K.; Wells, D.E.; Harland, R.M.; Zimmerman, L.B. (2009) **Rapid gynogenetic mapping of** *Xenopus tropicalis* mutations to chromosomes. Dev Dyn 238:1398–1346.

Kidwell, M. G.; Lisch, D. R. (2001). **Perspective: Transposable elements, parasitic DNA, and genome.** National Library Medicine. *55(1), 1–24.* doi:10.1111/j.0014-3820.2001.tb01268.x

Klein, S. J.; O'Neill, R. J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosome Research, –. doi:10.1007/s10577-017-9569-5

Lavie, L.; Maldener, E.; Brouha, B.; Meese, E.U.; Mayer, J. (2004) The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. Genome Res 14:2253–2260. https://doi.org/10.1101 /gr.2745804

Lexa, M.; Lapar, R.; Jedlicka, P.; Vanat, I.; Cervennansky, M.; Kejnovsky, E. (2018). **TEnester: a recursive software tool for structure-based discovery of nested transposable elements.** IEEE International Conference on Bioinformatics and Biomedicine (BIBM) - 2776–2778. doi:10.1109/bibm.2018.8621071

Mable, B. K; Alexandrou, M. A; Taylor, M. I (2011). **Genome duplication in amphibians** and fish: an extended synthesis. *284(3)*, *151–182*. doi:10.1111/j.1469-7998.00829.x

Marracci, S.; Batistoni, R.; Pesole, G.; Citti, L.; Nardi, I (1996). Gypsy/Ty3-like elements in the genome of the terrestrial salamander Hydromantes (Amphibia, Urodela). 43(6), 584–593. doi:10.1007/bf02202106

Matsuda, Y.; Uno, M.; Kondo, M.J.; Gilchrist, A.M.; Zorn, D.S.; Rokhsar, M.; Schmid, M. (2015). **A New Nomenclature of** *Xenopus laevis* **Chromosomes Based on the Phylogenetic Relationship to** *SiluranalXenopus tropicalis.* **Cytogenet. Genome Res., 145, pp. 187-191**

McLaughlin, R.N. Jr.; Malik, H.S. (2017) Genetic conflicts: the usual suspects and beyond. J Exp Biol 220:6–17. https://doi.org/10.1242/jeb.148148

McVean, G. (2010) What drives recombination hotspots to repeat DNA in humans? Philos Trans R Soc Lond B Biol Sci. 365: 1213-1218.

Mezzasalma, M.; Glaw, F.; Odierna, G.; Petraccioli, A.; Guarino, F.M. (2015) **Karyological analyses of Pseudhymenochirus merlini and Hymenochirus boettgeri provide new insights into the chromosome evolution in the anuran family Pipidae.** Zool. Anz. 258, 47–53.

Muotri, A.R; Marchetto, M. C.N.; Coufal, N. G.; Gage, F. H. (2007) **The necessary junk: new functions for transposable elements.** *Human Molecular Genetics*, Volume 16, Issue R2, Pages R159–R167, <u>https://doi.org/10.1093/hmg/ddm196</u>

Park, H.J.; Gokhale, C.S.; Bertels, F. (2021) **How sequence populations persist inside bacterial genomes.** Genetics 217, iyab027. Doi:10.1093/genetics/ iyab027

Peterson, D. G. (2002). Integration of Cot Analysis, DNA Cloning, and High-Throughput Sequencing Facilitates Genome Characterization and Gene Discovery. Genome Research, *12(5)*, *795–807*. doi:10.1101/gr.226102

Petrov, D.A.; Fiston-Lavier, A-S; Lipatov, M.; Lenkov, K.; Gonzalez, J. (2011) **Population** genomics of transposable elements in *Drosophila melanogaster*. Mol Biol Evol.28(5):1633–44. Phillips, C.M.; Brown, K.C.; Montgomery, B.E.; Ruvkun, G.; Montgomery, T.A. (2015) piRNAs and piRNA-dependent siRNAs protect conserved and essential *C. elegans* genes from misrouting into the RNAi pathway. Dev Cell 34:457–465. https://doi.org/10.1016/j.devcel.2015.07.009

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.

Rishishwar, L.; Tellez Villa, C.E.; Jordan, I.K. (2015) **Transposable element polymorphisms recapitulate human evolution**. Mob DNA. 6:21.

Roco, A. S.; Liehr, T.; Ruiz-García, A.; Guzmán. K.; Bullejos, M. (2021) Comparative Distribution of Repetitive Sequences in the Karyotypes of *Xenopus tropicalis* and *Xenopus laevis* (Anura, Pipidae). Genes (Basel) Apr 21;12(5):617. doi: 10.3390/genes12050617.

Roco, Á. S.; Olmstead, A. W.; Degitz, S. J.; Amano, T.; Zimmerman, L. B.; Bullejos, M. (2015). Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis.* Proceedings of the National Academy of Sciences, 112(34), E4752–E4761. doi:10.1073/pnas.1505291112

Rogers, R. L., Zhou, L., Chu, C., Márquez, R., Corl, A., Linderoth, T., ... Nielsen, R. (2014). Genomic takeover by transposable elements in the Strawberry poison frog. Molecular Biology and Evolution. doi:10.1093/molbev/msy185

Rogers, R.L. (2015) Chromosomal rearrangements as barriers to genetic homogenization between archaic and modern humans. Mol Biol Evol 32:3064–3078. https://doi.org/10.1093/molbev/msv204

Scalvenzi, T. ; Pollet, N (2014). **Insights on genome size evolution from a miniature inverted repeat transposon driving a satellite DNA.** Molecular Phylogenetics and Evolution, 81(), 1–9. doi:10.1016/j.ympev.2014.08.014 Schneider, K.L.; Xie, Z.; Wolfgruber, T.K.; Presting, G.G. (2016) **Inbreeding drives maize centromere evolution.** Proc Nat Acad Sci U S A 113:E987–E996. <u>https://doi.org/10.1073</u>/pnas.1522008113

Seifertova, E.; Zimmerman, L. B; Gilchrist, J. M.; Macha, J. (2013). Efficient highthroughput sequencing of a laser microdissected chromosome arm. BMC Genomics. 14(1), 357–. doi:10.1186/1471-2164-14-357

Senft, A. D.; e Macfarlan, T. S. (2021). **Transposable elements shape the evolution of mammalian development.** Nature Reviews Genetics. doi:10.1038/s41576-021-00385-1

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

Sinzelle, L.; Carradec, Q.; Paillard, E.; Bronchain, O. J.; Pollet, N. (2011). Characterization of a Xenopus tropicalis Endogenous Retrovirus with Developmental and Stress-Dependent Expression. Journal of Virology, 85(5), 2167–2179. doi:10.1128/JVI.01979-10

Siomi, M.C.; Sato, K.; Pezic, D.; Aravin, A.A. (2011) **PIWI-interacting small RNAs: the vanguard of genome defence.** Nat Rev Mol Cell Biol 12:246–258. <u>https://doi.org/10.1038/nrm3089</u>

Slotkin, R.K.; Martienssen, R. (2007) **Transposable elements and the epigenetic regulation of the genome.** Nat Rev Genet. 8(4):272–85.

Slotkin, R.K.; Martienssen, R. (2007) **Transposable elements and the epigenetic regulation of the genome.** Nat Rev Genet 8:272–285. <u>https://doi.org/10.1038/nrg2072</u>

Sun, C.; Shepard, D. B.; Chong, R. A.; Lopez Arriaza, J.; Hall, K.; Castoe, T. A.; Feschotte, C.; Pollock, D. D.; Mueller, R. L. (2012). **LTR Retrotransposons Contribute to Genomic Gigantism in Plethodontid Salamanders.** Genome Biology and Evolution, 4(2), 168–183. doi:10.1093/gbe/evr139

Tolomeo, D.; Capozzi, O.; Stanyon, R. R.; Archidiacono, N.; D'Addabbo, P.; Catacchio, C. R.; Purgato, S.; Perini, G.; Schempp, W.; Huddleston, J.; Malig, M.; Eichler, E. E.; Rocchi, M. (2017). **Epigenetic origin of evolutionary novel centromeres.** Scientific Reports, *7(), 41980–.* doi:10.1038/srep41980

Tymowska, J.; Fischberg, M. (1982) A comparison of the karyotype, constitutive heterochromatin, and nucleolar organizer regions of the new tetraploid species Xenopus epitropicalis Fischberg and Picard with those of Xenopus tropicalis Gray (Anura, Pipidae). Cytogenet. Cell Genet. 34, 149–157

Viana, P. F.; Ezaz, T.; de Bello Cioffi, M.; Liehr, T.; Al-Rikabi, A.; Goll, L. G.; Rocha, A. M.; Feldberg, E (2020). Landscape of snakeâ sex chromosomes evolution spanning 85 MYR reveals ancestry of sequences despite distinct evolutionary trajectories. Scientific Reports, 10(1), 12499–. doi:10.1038/s41598-020-69349-5

Vignali, R.; Nardi, I. (1996) **Unusual features of the urodele genome: do they have a role in evolution and development?.** Int J Dev Biol Aug;40(4):637-43.

Vogt, J.; Bengesser, K.; Claes, K. B. M.; Wimmer, K. (2014). **SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints.** Genome Biology, 15(6), –. doi:10.1186/gb-2014-15-6-r80

Wolf, G.; Greenberg, D.; Macfarlan, T.S. (2015) **Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Kruppel-associated box zinc finger protein family.** Mob DNA 6:17. https://doi.org/10.1186/s13100-015-0050-8

Wright, S.; Finnegan, D. (2001) **Genome evolution: sex and the transposable element.** Curr. Biol. royalsocietypublishing.org/journal/rstb Phil. Trans. R. Soc. B 377: 20200477 11 doi:10.1016/S0960-9822(01) 00168-3.

Zattera M.L., Lima L., Duarte I., de Sousa D.Y., Araújo O.G.D.S., Gazoni T., Mott T., Recco-Pimentel S.M., Bruschi D.P (2019) Chromosome Spreading of the (TTAGGG)n

Repeats in the Pipa Carvalhoi Miranda-Ribeiro, 1937 (Pipidae, Anura) Karyotype. *Comp. Cytogenet.* 13:297–309.

Zattera, M.L.; Gazolla, C.B.; de Soares, A.; Gazoni, T.; Pollet, N.; Recco-Pimentel, S.M.; Bruschi, D.P. (2020) **Evolutionary Dynamics of the Repetitive DNA in the Karyotypes of** *Pipa carvalhoi* **and** *Xenopus tropicalis* (Anura, Pipidae). Front. Genet. 11, 1–10.

Zhang, Y.; Li, S.; Abyzov, A.; Gerstein, M.B. (2017) Landscape and variation of novel retroduplications in 26 human populations. PLoS Comput Biol 13:e1005567. https://doi.org/10.1371 /journal.pcbi.1005567

ANEXO I

- Acessar o prompt de comando do Windowns. Ir para o diretório da pasta Bin.
 C:\Windows\system32:cd\ (enter)
 C:\>cd blast\bin (enter)
- 2- Formatar banco (genoma):

C:\Blast\bin\formatdb -i genome fna -p F -n genomeformated (enter)

3- Realizar o Blast

C:\Blast\bin\blastall -p blastn –i queryblastlocal –d genomeformated –o resultblastlocal (enter)

4- Filtrar os dados

Abrir o arquivo de resultado tabular. Filtrar os dados no excel. Escolher as sequências e ficar apenas com as colunas B (ID do contig/scaffold), I (posição inicial no contig/scaffold) e J (posição final no contig/scaffold). Salvar em bloco de notas com o nome "resultblastfiltrado".

5- Recuperar as sequências do transposon

Script de Gama et al. (2022) utilizando o Python para recuperar 2 kb de cada lado. Comando utilizado:

C:\pasta:python script2kb.pl