

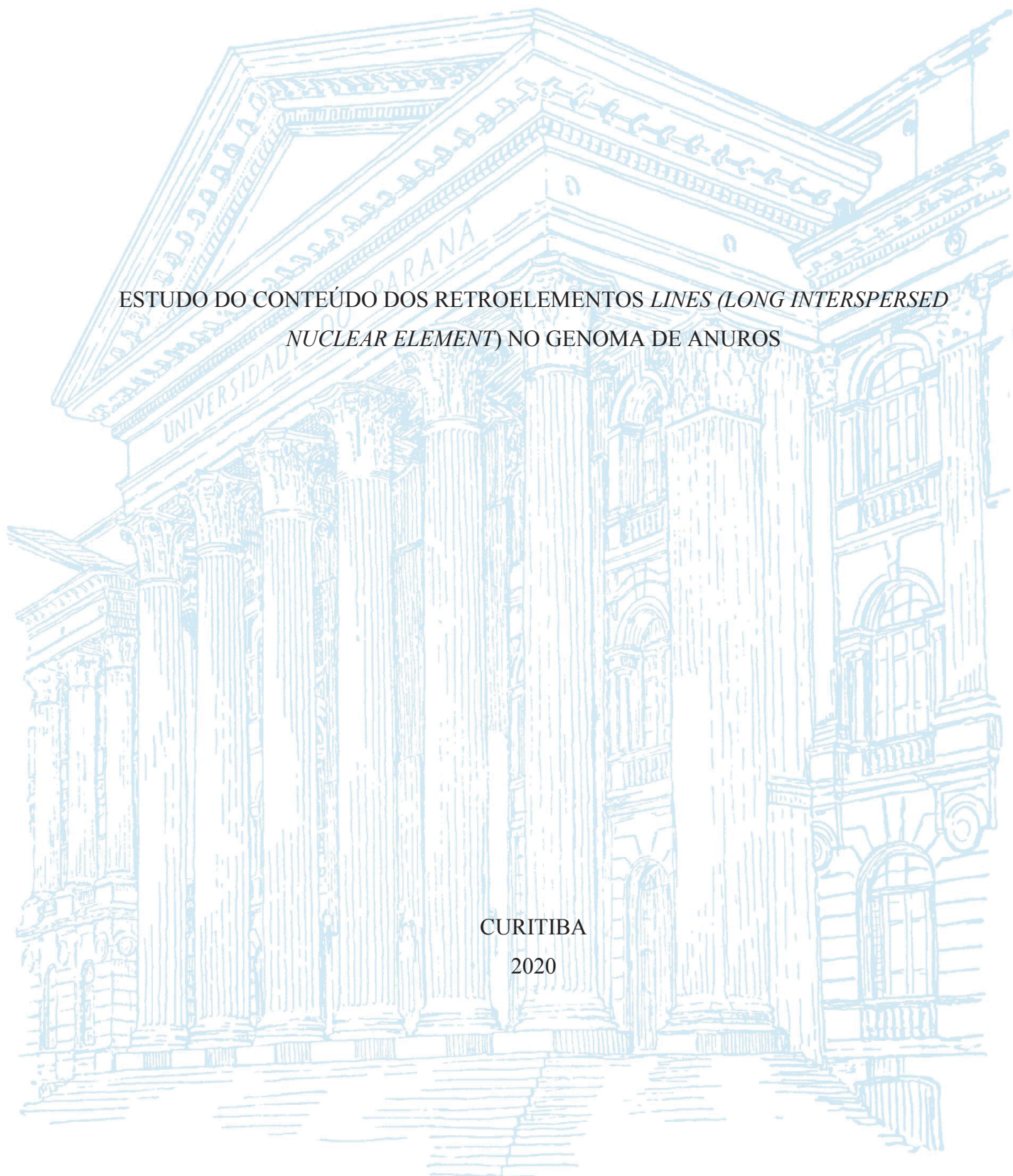
UNIVERSIDADE FEDERAL DO PARANÁ

JOANA DE MOURA GAMA

ESTUDO DO CONTEÚDO DOS RETROELEMENTOS *LINES* (*LONG INTERSPERSED
NUCLEAR ELEMENT*) NO GENOMA DE ANUROS

CURITIBA

2020



JOANA DE MOURA GAMA

ESTUDO DO CONTEÚDO DOS RETROELEMENTOS *LINES* (*LONG INTERSPERSED
NUCLEAR ELEMENT*) NO GENOMA DE ANUROS

Dissertação apresentada ao curso de Pós-Graduação em Genética, Setor de Biológicas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Genética.

Orientador: Prof. Dr. Daniel Pacheco Bruschi.
Coorientadora: Prof^a. Dr^a. Adriana Ludwig.

CURITIBA

2020

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Rosilei Vilas Boas – CRB/9-939).

Gama, Joana de Moura.

Estudo do conteúdo dos retroelementos Lines (Long Interspersed Nuclear Element) no genoma de anuros. – Curitiba, 2020.
98 f. : il.

Orientador: Daniel Pacheco Bruschi.

Coorientadora: Adriana Ludwig.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Anuros. 2. Retroelementos. 3. Genomas. 4. Anfíbio – Mata Atlântica. I. Título. II. Bruschi, Daniel Pacheco. III. Ludwig, Adriana. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 597.6



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 em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **JOANA DE MOURA GAMA** intitulada: **Estudo do conteúdo dos retroelementos Lines (Long Interspersed Nuclear Element) no genoma de anuros**, sob orientação do Prof. Dr. DANIEL PACHECO BRUSCHI, 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 mestre 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, 23 de Março de 2020.

Assinatura Eletrônica

03/04/2020 09:44:30.0

DANIEL PACHECO BRUSCHI

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

27/04/2020 09:19:20.0

LUCIANA BOLSONI LOURENÇO

Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica

03/04/2020 12:23:55.0

MARCELO RICARDO VICARI

Avaliador Interno (UNIVERSIDADE ESTADUAL DE PONTA GROSSA)

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: 39183

Para autenticar este documento/assinatura, acesse <https://www.pppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 39183

Dedico este trabalho ao meu pai Antônio Jorge (*in memoriam*), meu maior incentivador desde o início, alimentando minha curiosidade e vontade de ser cientista. Tenho certeza que você não perderia esse momento por nada.

AGRADECIMENTOS

Agradeço primeiramente a Fundação de Amparo à Pesquisa de São Paulo (FAPESP) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro durante o mestrado, em tempos de desmonte, eu fui privilegiada.

Ao Instituto Carlos Chagas (ICC/ Fiocruz Paraná), a Universidade Federal do Paraná (UFPR), ao Centro de Tecnologias Avançadas em Fluorescência (CTAF), por ceder a utilização dos microscópios para captura das imagens, ao departamento de Genética e seus respectivos laboratórios, com uma atenção especial aos laboratórios Citogenética Evolutiva e Conservação Animal (CECA- o qual faço parte), Laboratório de Citogenética Humana e Oncologia (LabCHO) e Laboratório de Interação Planta - Bactéria (LIPB), por ceder espaço, muitas vezes equipamentos de uso exclusivo de seu laboratório e sempre emprestando reagentes na hora da pressa.

Aos meus orientadores Daniel Pacheco Bruschi e Adriana Ludwig, pessoas especiais e pacientes, que sempre lembrarei com carinho. Seja pelos ensinamentos na área de genética, os quais foram muitos ou como amigos. Obrigado pela paciência e por terem me orientado tão bem, espero ser uma pesquisadora como vocês.

Agradecimento para “as meninas do Daniel” que se resume ao Laboratório Citogenética Evolutiva e Conservação Animal (CECA), como um grupo de pesquisa estamos sempre nos ajudando no possível. Meninas: Camilla, Michelle, Deborah, Julia, Gi Bueno, Amanda e Fernanda, vocês tornaram esses dois anos mais leves e divertidos. Agradecimento especial para Camilla Gazolla por tudo e principalmente por ter me ajudado em todas etapas dessa dissertação.

Aos meus amigos de Seropédica que me acompanham desde a graduação e me viram crescer profissionalmente, que mesmo longe ainda repartimos muitas coisas especiais. Em especial ao meu Professor e amigo Hélio Ricardo Silva, que contribuiu muito na minha formação como cientista e a entender algumas questões filosóficas da vida, e também aos amigos que conheci e cultivei nesses anos em Curitiba, valeu a pena ter vocês.

Um agradecimento mais que especial para minhas duas mães (Sandra Regina e Maria Augusta) que se preocupam e cuidam muito de mim, e me apoiam sempre nas minhas escolhas. Cada dia longe no final vai valer a pena. A minha família, principalmente as minhas tias que serem se orgulharam de qualquer detalhe conquistado.

Não poderiam faltar agradecimentos ao meu parceiro Pablo Silva, por me incentivar e me dar apoio nos mínimos detalhes, você faz grande diferença na minha vida, sempre trazendo mais alegria nos meus dias nublados.

“A História está repleta de pessoas que, como resultado do medo, ou por ignorância, ou por cobiça de poder, destruíram conhecimentos de imensurável valor que, em verdade, pertenciam a todos nós. Nós não devemos deixar isso acontecer de novo.”

Cosmos- Carl Sagan, 1980.

RESUMO

Elementos de transposição são sequências repetitivas com importante papel na evolução e funcionamento dos genomas. Existem poucos estudos que combinem Citogenética e Genômica (em citogenômica) em anuros, principalmente envolvendo essas sequências e seu mapeamento nos cromossomos. O objetivo desse trabalho foi estudar o conteúdo de retroelementos do tipo LINE no genoma de anuros, com enfoque na espécie *Pipa carvalhoi*, uma espécie endêmica da Mata Atlântica pertencente à família Pipidae. A presença dos elementos do tipo LINE, com foco especial nos elementos genericamente nomeados Rex (Rex1, Rex3, Rex5 e Rex 6), foi investigada por BLAST no genoma dos anuros e as cópias encontradas foram caracterizadas quanto a presença de fases abertas de leitura e domínios. O domínio da transcriptase reversa foi utilizado para acessar as relações evolutivas dessas sequências. A análise dessas sequências identificou quatro grupos distintos, que correspondem aos elementos classificados na literatura como Rex1 – grupo Rex1, Rex3 – grupo RTE, Rex5 – grupo L2 e Rex6 – grupo R4. De forma geral, foi possível ter um panorama da presença desses elementos em anuros, sendo que eles parecem ser muito antigos nos genomas de vertebrados, com diversidade de cópias já em espécies ancestrais, porém ondas de retrotransposição devem ter ocorrido mais recentemente, como para os elementos Rex1 e L2 de *P. carvalhoi*. Ainda, perdas secundárias independentes de alguns elementos devem ter acontecido em algumas espécies para explicar a sua distribuição. Um possível evento de transferência horizontal de um elemento do grupo RTE entre *P. carvalhoi* e *Helobdella robusta* (hirudínea) foi testado pela comparação da taxa de substituição sinônima das cópias do elemento e de genes constitutivos das espécies hospedeiras, indicando a não ocorrência desse evento de TH. Além disso, os elementos Rex1 e RTE foram mapeados nos cromossomos dos anuros *P. carvalhoi* e *Xenopus tropicalis*, pela primeira vez, através da técnica de *Hibridação in situ fluorescente*. Sinais de hibridação do elemento Rex1 foram detectados apenas nos cromossomos 3, 6, 7 e 10 de *P. carvalhoi*, em regiões heterocromáticas e associados com motivos de microssatélites (CA)₁₅, (GA)₁₅, (GATA)₈, (CGC)₁₀ e (CAG)₁₀, já reportados em outros trabalho. Nesse trabalho, foram estudados pela primeira vez o conteúdo de retroelementos LINEs no genoma de anuros, gerando informações básicas sobre a distribuição e evolução desses elementos que podem contribuir para o entendimento do papel dessas sequências na evolução e diversificação desses genomas.

Palavras-chaves: *LINE, Anura, Rex 1, RTE.*

ABSTRACT

Transposable elements are repetitive sequences with an important role in the evolution and functioning of genomes and compose about 30% of the anuran genomes. There are few cytogenomics studies in anurans, specially involving these sequences and their physical mapping on chromosomes. The goal of this work was to study the content of LINE retroelements in the anuran genome, mainly in *Pipa carvalhoi*, an endemic species of the Atlantic Forest belonging to the Pipidae family. The presence of LINE elements, with focus in those generically named Rex (Rex1, Rex3, Rex5 and Rex 6), was investigated with BLAST searches in the anuran genomes available. The copies were characterized for the presence of open reading frames and domains. The domain of reverse transcriptase was used to access the phylogenetic relationships of these sequences. Our results corroborate the idea that elements called Rex belong to different clades/superfamilies (Rex1 - clade Rex1, Rex3 - clade RTE, Rex5 - clade L2 and Rex6 - clade R4). In general, it was possible to obtain an overview of the presence of these elements in anurans. They appear to be very old in the vertebrate genomes, with a diversity of copies already existing in ancestral species; however, waves of retrotransposition also occurred more recently, as suggested for the elements Rex1 and L2 of *P. carvalhoi*. Also, independent secondary losses of some elements must have occurred in some lineages to explain their distribution. A possible event of horizontal transfer of an RTE element between *P. carvalhoi* and *Helobdella robusta* (Hirudinea) was tested by comparing the rate of synonymous substitution of the RTE copies and host species constitutive genes, indicating was not possible a horizontal transfer of this element. Also, Rex1 and RTE were mapped on the chromosomes of the *P. carvalhoi* and *X. tropicalis* for the first time, using the fluorescent in situ hybridization technique. Hybridization signals for the Rex1 element were detected only on chromosomes 3, 6, 7 and 10 of *P. carvalhoi*, in heterochromatic regions and associated with microsatellite motifs (CA)₁₅, (GA)₁₅, (GATA)₈, (CGC)₁₀ and (CAG)₁₀. In this work, the content of LINE retroelements was studied in the anuran genome for the first time, generating basic information about the distribution and evolution of these elements contributing to the understanding of the role of these sequences in the evolution and diversification of these genomes.

Keywords: LINE, Anura, Rex 1, RTE, Horizontal Transfer

LISTA DE FIGURAS

CAPÍTULO 1

- FIGURA 1 – Árvore filogenética reconstruída por máxima verossimilhança (ML), modelo LG+G, a partir do domínio da RT dos elementos LINEs recuperados no genoma de *P. carvalhoi*. Os valores de bootstrap são indicados próximo aos nós. Cada clado está representado por uma cor assim como cada espécie de anuro..... 42
- FIGURA 2 – Cladograma contendo as espécies de anuros com genomas sequenciados e incluídas em nossas análises ilustrando os dados de presença/ausência das principais subfamílias/clados de TEs da ordem LINE (ver legenda). As relações evolutivas entre as espécies foram compiladas a partir da plataforma Timetree (<http://www.timetree.org/book>) para a construção do cladograma. Círculos preenchidos representam nós inferidos com alta capacidade de refletir o real relacionamento filogenético entre os grupos, enquanto círculos não-preenchidos representam nós cujo relacionamento entre as espécies foram criados artificialmente durante a resolução das politomias observadas, de acordo com Hedges et al (2015). 44
- FIGURA 3 – Estimativa da taxa de substituições sinônimas (dS) para genes constitutivos, conforme listado na tabela 4 e do elemento RTE (pipa84207 e RTE5_HRo). 47

CAPÍTULO 2

- FIGURA 1 – Representação de regiões genômicas contendo cópias do do Retroelemento Rex1. A) Rex1 de *P. carvalhoi* desenhado com base no contig pipa381153. B) Rex1 de *X. tropicalis* desenhado com base na sequência presente no cromossomo 5. As linhas pontilhadas em rosa representam a região cromossômicas, as ORFs estão representadas de uma mesma cor (conforme legenda) e seus domínios de cores mais escuras. A régua acima ilustra o tamanho do contig/região cromossômica isolada, evidenciando as posições das ORFs e domínios. Na figura 1B está destacado o final do elemento contendo a cauda Poly-A..... 78
- FIGURA 2 – Representação esquemática da região genômica contendo cópia do LINE RTE de *P. carvalhoi*, baseado no contig pipa84507. As linhas pontilhadas (em

verde escuro) representam a região cromossômica, as ORFs estão representadas de uma mesma cor (conforme a legenda) e seus domínios de cores mais escuras. A régua acima ilustra o tamanho do contig isolado evidenciando as posições das ORFs e domínios. 79

FIGURA 3 – : Mapeamento dos elementos Rex 1 nos cromossomos de *P. carvalhoi*. (A) e *X. tropicalis* (B). As setas evidenciam os sinais considerados como marcação do elemento na figura A. 80

FIGURA 4 – . Cromossomos de *P. carvalhoi*, as imagens consecutivamente representam: coloração em DAPI, marcação inespecífica da sonda marcada com 5-TAMRA-dUTP e merge das duas imagens. 83

LISTA DE TABELAS

CAPÍTULO 1

TABELA 1 – Informações dos genomas de anuros disponíveis na plataforma do Genbank utilizados nesse trabalho.....	33
TABELA 2 – Genes ortólogos em <i>X. tropicalis</i> e <i>H. robusta</i> usados para calcular a taxa dS e testar a hipótese de TH.	36
TABELA 3 – Classe, ordem e famílias dos elementos de transposição identificados no genoma de <i>Pipa carvalhoi</i> , através da pipeline PIRATE. A classificação desses elementos é baseada em Wicker et al (2007).	37
TABELA 4– Queries do Repbase utilizadas no tBLASTn contra o genoma de <i>P. carvalhoi</i> com os respectivos clados formados pelas sequências recuperadas desse genoma.	43
TABELA 5 – Valores estimados de número efetivo de códons (ENC) e índice de códon bias (CBI), calculados no DAMBE para cada gene constitutivo e do RTE comparando <i>X. tropicalis</i> e <i>H. robusta</i>	46

CAPÍTULO 2

TABELA 1 – Sequência dos primers utilizados nesse trabalho com informações dos contigs utilizados em sua construção, a temperatura de anelamento (°C) e o tamanho dos fragmentos estimados em pares de bases (pb).	75
TABELA 2 – : Resumo do resultado das buscas por Blastn de cópias do elemento Rex1 no genoma de <i>X. tropicalis</i> . São apresentados os hits com alinhamento maior de 100 pb.....	82

LISTA DE ABREVIATURAS OU SIGLAS

- BA -Análise Bayesiana
- CBI - índice de códon *bias*
- DAMBE *Data Analysis in Molecular Biology and Evolution*
- DNA -Ácido desoxirribonucleico
- dS -Substituição Sinônima
- dUTP - Deoxyuridine triphosphate
- ENC - Número efetivo de códons
- EN - Endonuclease
- FISH - Hibridização *in situ* Fluorescente
- Gpb - Giga pares de bases
- LINE - *Long Interspersed Nuclear Elements*
- MCMC Markov Monte Carlo
- MITE *Miniature Inverted-repeat Transposable Elements*
- ML - Máxima Verossimilhança
- μl -Microlitros
- noCat - Não categorizadas
- ng/μl - Nanogramas por microlitros
- ORF -*Open read frame* Endonuclease
- pb - Pares de bases
- PBS Tampão fosfato-salino
- PCR -Reação em Cadeia da Polimerase Endonuclease
- pmol -PicoMoles
- RT -Transcriptase Reversa
- SSR - Microssatélites

TE Elemento de transposição

TH Transferência Horizontal

TPRT *Target-Primed Reverse Transcription*

TRIM *Terminal-repeat Retrotransposons In Miniature*

TV Transferência Vertical

U Unidades

SUMÁRIO

1	INTRODUÇÃO	14
2	OBJETIVOS	26
2.1	OBJETIVOS GERAIS	26
2.2.	OBJETIVOS ESPECÍFICOS	27
CAPÍTULO 1		28
1.	RESUMO	28
2.	INTRODUÇÃO.....	28
3.	MATERIAL E MÉTODOS.....	31
4.	RESULTADOS E DISCUSSÃO	36
5.	ANEXOS	51
CAPÍTULO 2		72
1.	RESUMO	72
2.	INTRODUÇÃO.....	72
3.	MATERIAL E MÉTODOS.....	74
4.	RESULTADOS E DISCUSSÃO	76
5.	ANEXOS	85
3.	CONCLUSÕES E CONSIDERAÇÕES FINAIS	91
REFERÊNCIAS		92

1. INTRODUÇÃO

Elementos de Transposição

Os elementos transponíveis (TEs) são sequências móveis de DNA com capacidade de se transpor de um lugar para outro. Estão dispersos no genoma e são altamente repetidos, sendo encontrados em todas as espécies eucarióticas, podendo constituir aproximadamente 45% do genoma humano, 80% do genoma de algumas plantas e entre 3-20% em fungos (BIÉMONT; VIEIRA, 2006; WICKER *et al.*, 2007).

Variedade na densidade e composição de TEs é observada entre organismos eucariotos, como por exemplo, o genoma de *Aedes aegypti*, *Homo sapiens*, *Mus musculus*, *Saccharomyces cerevisiae* e *Schizosaccharomyces pombe*, onde existe predominância de TEs que necessitam de uma transcriptase reversa (RT) para se transpor, esses elementos são os principais responsáveis pelo aumento desses genomas (FESCHOTTE; PRITHAM, 2007; SUN *et al.*, 2015). Já em *Anopheles gambiae*, *Caenorhaditis elegans*, *Entamoeba invadens*, *Oryza sativa* e *Trichomonas vaginalis*, existe prevalência de transposons que necessitam apenas de uma transposase. (FESCHOTTE; PRITHAM, 2007).

Classificar elementos de transposição é um trabalho complexo e árduo, necessitando de uma análise criteriosa das sequências desses elementos e sua comparação por similaridade com outras sequências previamente descritas (KAPITONOV; JURKA, 2008; WICKER *et al.*, 2007). Em 1989, Finnegan sugeriu uma classificação dos TEs baseada em seu mecanismo de transposição, separando-os em elementos de classe I e II. Assim, os elementos de Classe I necessitam de um RNA intermediário e uma transcriptase reversa (RT), para produzir uma nova molécula. Essa nova fita é inserida em um novo local no genoma através de uma integrase (IN) ou endonuclease (EN) (FINNEGAN, 1989). Já os elementos de Classe II realizam transposição “direta”, através de uma molécula de DNA (FINNEGAN, 1989), conhecidos como transposon de DNA e que sua transposição é mediada por uma enzima transposase (WICKER *et al.*, 2007).

Inicialmente os elementos de classe I foram separados como *Retrovirus-like elements* e *Nonviral retroposons* (FINNEGAN, 1989), mais conhecidos posteriormente como LTR e Não-LTR Retrotransposons (CAPY *et al.* 1997). Os primeiros possuem longas repetições terminais (LTRs) diretas, e uma região interna contendo *Open Reading Frames* (ORFs)

similares aos genes de retrovírus denominados *gag*, que codifica uma proteína com domínio similar ao do capsídeo viral, e *pol* que resulta em uma poliproteína portadora dos domínios RT, RNaseH, IN de uma protease aspártica. Em alguns elementos essas ORFs são fusionadas e em outros uma terceira ORF similar ao gene viral *env* pode ser encontrada. Os Não-LTRs são os elementos sem repetições terminais e em geral possuem duas ORFs, a primeira similar ao gene *gag* e a segunda correspondente a uma RT e uma EN e uma cauda rica em A no final da extremidade 3' (FINNEGAN, 1989).

De acordo com Finnegan (1989), os TEs de classe II podem ser separados em elementos com terminações repetidas invertidas curtas e longas (hoje classificados como elementos da superfamília *Mutator*). Atualmente, não existe mais essa distinção e todos os elementos contendo TIRs (Terminal Inverted Repeat) estão em um mesmo subclasses de transposons de DNA (ver abaixo).

Nesse modelo de classificação, ainda não estavam contemplados diversos elementos distintos que não se enquadravam em nenhuma dessas classes, como o elemento de *Dictyostelium discoideum* (DIRS) (FINNEGAN, 1989) e elementos do tipo Crypton, Maverick e Helitrons.

A denominação das duas classes cunhada por Finnegan (1989) (elementos de classe I e II) perdura até os dias de hoje e vem sendo continuamente sendo refinada. Uma proposta unificada de classificação de TEs foi realizada por Wicker *et al.* (2007) e leva em consideração a estrutura do elemento, relações entre as sequências e quais enzimas participam do processo de replicação. No entanto, a utilização do sistema de classificação de Wicker *et al.* (2007) não é unanimidade entre os pesquisadores (KAPITONOV; JURKA, 2008; SEBERG; PETERSEN, 2009) e com frequência é possível encontrar algumas discrepâncias entre fontes distintas de informação.

De acordo com a classificação de Wicker *et al.* (2007), a classe I, popularmente conhecida como “copia e cola”, pode ser dividida em cinco ordens: LTR (*Long Terminal Repeat*), DIRS (*Dictyostelium Intermediate Repeat Sequence*), LINE (*Long Interspersed Nuclear Element*), SINE (*Short Interspersed Nuclear Element*) e PLE (*Penelope-Like*) (WICKER *et al.*, 2007).

Os elementos de classe II (genericamente conhecidos como Transposons) são divididos atualmente em duas subclasses (I e II), segundo Wicker *et al.* (2007). A primeira é

popularmente conhecida por possuir mecanismo de transposição “corta e cola” e apresenta duas ordens: TIR (*Terminal inverted repeat*) e Crypton. Elementos da ordem TIR possuem as extremidades invertidas (TIRs) e codificam uma enzima chamada transposase, a enzima responsável pela transposição nessa ordem, e a qual reconhece as TIRs, corta e cola o elemento em outro local no genoma. São reconhecidas nove superfamílias de elementos TIRs: *CACTA*, *hAT*, *Merlin*, *Mutator*, *P*, *PIF-Harbinger*, *PiggyBac*, *Tc-1-Mariner* e *Transib*. De acordo com o Repbase (KAPITONOV; JURKA, 2008), um banco de dados de sequências repetitivas, os elementos TIRs podem ser divididos em um número maior de superfamílias. Já a ordem Crypton é composta de apenas uma superfamília, a *Crypton* (WICKER *et al.*, 2007), que reúne os elementos que codificam uma tirosina recombinase ao invés de uma transposase.

A subclasse II é dividida em duas ordens: *Helitron* e *Maverick*, com superfamílias correspondentes ao nome da ordem e utilizam o mecanismo de transposição conhecido como replicação via círculo rolante de apenas uma das fitas de DNA (FESCHOTTE; PRITHAM, 2007; WICKER *et al.*, 2007).

Um tipo específico de transposons de DNA curtos são os MITES, denominados *Miniature Inverted-repeat Transposable Elements*, um subconjunto de transposons não autônomos. Esses elementos só podem ser mobilizados pelas transposases de elementos completos e podem se originar de uma descaracterização de qualquer elemento de Classe II da ordem TIRs, como os da superfamília *Tc1/Mariner* ou *hAT* (JIANG *et al.*, 2004).

Durante a inserção da maioria dos elementos de classe I e II, são geradas duplicações do sítio alvo (TSDs - *Target Site Duplication*), que funcionam como uma assinatura molecular do evento de transposição. Essa assinatura molecular tem sido utilizada como critério para distinguir os elementos em distintas superfamílias e o nível de conservação pode ser utilizado em inferências sobre o tempo de inserção de TE em um determinado cromossomo (WICKER *et al.*, 2007).

Outra classificação bastante difundida é a utilizada pelo REPBASE (<https://www.girinst.org/replibase/>), um importante repositório para sequências repetitivas apresenta pequenas diferenças, como é o caso da classificação das sequências depositadas nesse banco de dados (KAPITONOV; JURKA, 2008). Essa classificação apresenta de forma geral um número maior de superfamílias em relação ao trabalho de Wicker *et al.* (2007) e os dois grupamentos originais de retrotransposons (LTR e não-LTR) são mantidos. São

estabelecidas sete superfamílias dentro dos elementos LTRs (Gypsy, Copia, BEL, DIRS e as superfamílias de retrovírus endógenos ERV1, ERV2 e ERV3), enquanto os elementos não-LTRs são divididos em 16 superfamílias, das quais podemos concluir que 12 são elementos do tipo LINE (CR1, CRE, I, Jockey, L1, NeSL, R2, R4, RandI, Rex1, RTE e Tx1), de três superfamílias de elementos não autônomos (SINE1, SINE 2 e SINE 3), além da superfamília Penelope (KAPITONOV; JURKA, 2008). Os transposons de DNA são divididos em 3 grupos, sendo eles, recorta-e-cola, DNA de círculo-rolante (*Helitrons*) e os transposons auto-sintetizantes (*Politrans* ou *Maverick*), refletindo três mecanismos diferentes de transposição. As duas últimas classes são representadas por apenas uma superfamília, a qual leva o nome das próprias classes, já a classe recorta-e-cola é composta por 15 superfamílias: *Chapaev*, *En/Spm (CACTA)*, *hAT*, *Harbinger (Pif)*, *ISL2EU(IS4EU)*, *Kolobok*, *Mariner*, *Merlin*, *Mirage*, *MuDR (MULE)*, *Novosib*, *P*, *PiggyBac*, *Rehavkus* e *Transib* (KAPITONOV; JURKA, 2008).

De forma geral, a classificação feita por Wicker *et al.* (2007) dividindo os retrotransposons em cinco ordens no lugar de apenas dois grupos, assim como a separação dos elementos de classe II. Porém, a divisão em superfamílias parece não estar de acordo com a diversidade dos elementos, principalmente para elementos TIR e LINE e a classificação de Kapitonov e Jurka (2008) parece mais adequada em nível de superfamília.

Sistemas de classificação, como é de se esperar, não são completamente isentos de inconsistências, o que claramente podemos perceber quando tentamos categorizar os elementos genéticos móveis. Em parte, devido à rápida taxa de divergência entre as sequências, o que faz com que assinaturas moleculares sejam apagadas e/ou difíceis de serem categorizadas. Outro fator determinante é a grande diversidade de TEs que eventualmente vêm sendo descritos nos últimos anos, principalmente em decorrência do maior número de genomas sequenciados, revelando que alguns elementos são muito distintos dos já encontrados e não se enquadram nas classificações atuais.

DNA lixo? Parece que não.

Desde a descoberta dos TEs por Barbara McClintock em 1940, que observou a dinâmica dos elementos denominados *Ds* e *Ac*, por meio de quebras no braço curto do cromossomo 9 do milho. Ela pode observar que esses elementos poderiam mudar sua posição dentro do cromossomo de forma aleatória, descobrindo assim sequências que poderiam se

movimentar no genoma. Anteriormente não se conhecia a função e nem a importância das sequências repetitivas nos genomas e assim pensava-se que os TEs, como outras sequências repetitivas, eram DNA lixo ou DNA “egoísta”. No entanto com o aumento dos estudos dessas sequências começaram a entender como elas agem e que são substratos genômicos importantes na diversidade genética e evolução (BIÉMONT, 2010; BIÉMONT; VIEIRA, 2006).

A capacidade de mobilidade dos elementos transponíveis influencia diretamente na arquitetura do genoma, podendo com o tempo levar a mudanças genômicas em grande escala, como seu envolvimento em rearranjos cromossômicos, através da recombinação ectópica (JURKA *et al.*, 2007; WICKER *et al.*, 2007) ou em microescala, como na formação de novos genes (PARKER *et al.*, 2009) e/ou alteração de funções (BIESSMANN *et al.*, 1990; PARDUE *et al.*, 2005). Por exemplo, em cães, a inserção de um elemento *LINE* no promotor do gene funcional *fgf4* causa fenótipo de pernas curtas (fenótipo condrodisplásico) quando em homozigose. Esse elemento provavelmente forneceu os componentes reguladores necessários para promover a expressão do gene *fgf4*, uma vez que esse gene está presente em cães não condrodisplásico, mas não é expresso (PARKER *et al.*, 2009).

Em *Drosophila*, a manutenção das extremidades dos cromossomos não é realizada pela telomerase e sim por dois transposons *LINEs*, *Het-A* e *TART*, que se inserem nas extremidades dos cromossomos, compensando a perda de nucleotídeos terminais durante a replicação do DNA (SILVA-SOUSA; CASACUBERTA, 2012). Essas cópias são mais complexas que as repetições inseridas pela telomerase, no entanto, possuem semelhanças com os telômeros comumente encontrados (BIESSMANN *et al.*, 1990; PARDUE *et al.*, 2005).

Ao que se refere a sua participação na diversificação dos cariótipos, os elementos de transposição podem estar direta ou indiretamente relacionados aos rearranjos cromossômicos, responsáveis pela evolução cromossômica (FARRÉ *et al.*, 2011; GRAY, 2000). Alguns autores sugerem que grande parte da diversidade cariotípica é causada pelos rearranjos, que geralmente ocorrem através de pontos de rupturas evolutivos (*Evolutionary Breakpoints-EBRs*) (FARRÉ *et al.*, 2011; LARKIN *et al.*, 2006). Durante a evolução, há recorrência de quebras nos mesmos pontos, inferindo que esses EBRs são reutilizados e frequentemente apresentam como característica comum o fato de estarem próximos a regiões de DNA repetitivo, como TEs e repetições *in tandem* (FARRÉ *et al.*, 2011; GLUGOSKI *et al.*, 2018;

LARKIN *et al.*, 2006). As repetições *in tandem* são classificadas em microssatélites, minissatélites ou satélites, frequentemente são ricas em repetições palindrômicas, formando grampos e facilitando que o DNA adquira uma conformação diferente do DNA-B, proposto por Watson & Crick, 1953 (LEACH, 1994). Essa conformação não-canônica pode favorecer a ocorrência de rupturas nas fitas de DNA, promovendo instabilidade do genoma, o que poderia levar a mudanças cromossômicas evolutivas (FARRÉ *et al.*, 2011).

Frequentemente tem sido reportada a presença de TEs associados com blocos de repetições microssatélites nos genomas (KEJNOVSKY *et al.*, 2013; PRIZON *et al.*, 2018). Sequências de repetições curtas e ricas em adenina e timina (AT) podem favorecer a formação de estrutura secundárias de segunda-ordem na molécula de DNA (grampos e alças) que parecem favorecer a inserção de TE (KEJNOVSKY *et al.*, 2013) e ambos estão relacionados aos EBRs (FARRÉ *et al.*, 2011). Um exemplo disso pode ser encontrado nas EBRs de mamíferos recuperadas ao longo da evolução do grupo, principalmente nos grandes primatas (FARRÉ *et al.*, 2011). Foi encontrada uma associação dos elementos *Alu* (classe I, ordem SINE) e *LI* (classe I, ordem LINE) com repetições microssatélites. Uma explicação para tal associação seria que os microssatélites serviriam de região alvo para a inserção desses retrotransposons, pois se assemelham muito aos locais de clivagem canônica para esses elementos, devido à presença dessas cópias semelhantes e repetitivas ocorre o pareamento não-alélico, levando eventualmente a reorganização genômica por recombinação homóloga não-alélica (FARRÉ *et al.*, 2011; GRAY, 2000).

Alternativamente, a presença do microssatélite AAAT em certas regiões pode derivar de elementos transponíveis degenerados já inseridos no genoma, pois elementos *LI* e *Alu* possuem uma cauda rica em poli (A) e uma região repetida interna *in tandem* (FARRÉ *et al.*, 2011). Apesar de ser difícil distinguir se a associação entre elementos de transposição e repetições microssatélites é uma causa da degeneração do elemento de transposição ou um efeito da própria inserção desses elementos em uma determinada região genômica (BUSCHIAZZO; GEMMELL, 2006; WILDER; HOLLOCHER, 2001) essa observação tem bastante relevância na dinâmica evolutiva dessas sequências e de seus impactos na arquitetura do genoma (BUSCHIAZZO; GEMMELL, 2006).

Em muitos casos, os transposons estão relacionados de maneira indireta com a ocorrência de rearranjos, não sendo a sua transposição em si o mecanismo causador, mas sim,

sua natureza repetitiva que favorece a ocorrência de recombinações ectópicas (GRAY, 2000). Outros exemplos de TE que podem influenciar em rearranjos cromossômicos podem ser vistos no gênero *Drosophila*, em especial as superfamílias *P*, *hobo*, *Doc*, *roo*, *FB*, *BEL* e *I* (LIM; SIMMONS, 1994). O transposon *Galileo* (classe II, superfamília *P*) está envolvido em pontos de inversões dos cromossomos de *Drosophila buzzatti* (DELPRAT *et al.*, 2009). Essas inversões alteram a ordem dos genes, podendo ou não alterar seu conteúdo e ocorrerem nas EBR's por recombinação homóloga não alélica. Muitas cópias de TE acumulam nessas regiões se tornando *hotspots* secundários de inserções de TE (DELPRAT *et al.*, 2009). Em *D. melanogaster* foi visto que o elemento *hobo* é responsável pelos rearranjos no cromossomo X, uma vez que todas as deleções e inversões detectadas nesse cromossomo estavam associadas a esse elemento. Curiosamente, os elementos que causavam as deleções estavam na mesma orientação no cromossomo e os que estavam envolvidos em inversões estavam em orientação oposta, sugerindo que esses rearranjos são formados por recombinação ectópica (LIM; SIMMONS, 1994).

TEs também facilitam a formação de rearranjos em fungos. Elementos das subfamílias *Gypsy* e *Copia*, encontrados em todos eucariotos, são mediadores de rearranjos no genoma de *Saccharomyces cerevisiae* por meio de da recombinação e, segundo os autores, há uma preferência no sítio de inserção dos elementos da subfamília *Copia* no genoma de *S. cerevisiae*, ocorrendo principalmente naqueles que são genes transcritos pela RNA Polimerase III (KIM *et al.*, 1998). No trigo existem evidências de que um TE da superfamília CACTA (classe II) permaneceu em uma EBRs, mediando a formação do complexo *R-r*, que são genes responsáveis pela pigmentação da aleurona, coleóptilo, ponta das folhas, raízes e anteras (WALKERL *et al.*, 1995).

Portanto, os elementos transponíveis podem conferir dinâmica aos cariótipos, através de inversões, deleções, translocações, sendo uma fonte de variação nas populações e podendo contribuir, dada a dinâmica populacional, para uma possível especiação (BENNETZEN, 2005).

Elementos de transposição em Anuros

Atualmente são reconhecidas cerca de 7168 espécies de anuros (FROST, 2020). Dessas, apenas 13 têm seu genoma sequenciado e disponível para uso na plataforma do GenBank, sendo elas *Nanorana parkeri*, *Pyxicephalus adspersus*, *Rana catesbeiana* (*Lithobates catesbeianus*), *Rhinella marina*, *Xenopus tropicalis*, *X. laevis*, *Leptobrachium leishanense*, *Spea multiplicata*, *S. bombifrons*, *Scaphiopus holbrookii*, *Scaphiophu couchii*, *R. temporaria*, *Oophaga pumilio* (GENBANK, 2020). Baseado em dados de *X. tropicalis*, estima-se que aproximadamente um terço do genoma dos anuros é composto por elementos transponíveis, que em sua grande maioria são transposons de DNA (HELLSTEN *et al.*, 2010). No entanto, os estudos desse conteúdo genômico em anfíbios ainda são escassos, considerando o grande acúmulo de sequências de natureza repetitiva descrita para esses poucos genomas sequenciados.

Sun e colaboradores (2015) sugerem que a distribuição e conservação de TEs em anfíbios mostra diferenças em relação a observada em mamíferos e aves. Como já mencionado, aproximadamente 30% do genoma de *X. tropicalis* é composto por TEs, enquanto os TEs compõem 9% do genoma das aves e 40-50% do genoma de mamíferos. Também tem sido reportado que, dentre as cópias descritas de TEs em anfíbios, grande parte estão integras e potencialmente inativos, um panorama muito semelhante ao observado em peixes, insetos e plantas e, oposto daquele observado em mamíferos e aves (HELLSTEN *et al.* 2010). Outra questão bastante interessante é que elementos como *Copia*, *BEL* e *Gypsy* não são encontrados nos genomas de aves e mamíferos, sugerindo que esses elementos podem ter se tornado imóveis depois da divergência da linhagem de anfíbios (HELLSTEN *et al.*, 2010; SUN *et al.*, 2015) e, portanto degenerados durante a evolução.

Chalopin *et al.* (2015) compara os padrões de diversidade de elementos transponíveis entre várias classes de vertebrados, reconhecendo a presença de 21 superfamílias em *X. tropicalis*, o segundo genoma mais diverso em superfamílias de TE dentre os vertebrados analisados, sendo superado apenas pelos peixes pertencentes a classe dos Actinopterygii, os quais possuem 26 superfamílias de TEs. O genoma de *X. tropicalis* é rico em elementos transponíveis de DNA, onde predominam as superfamílias *Tc1-Mariner*, *hAT*, *Harbinger*, *PiggyBac* e *T2/Kolobok*, seguido pela predominância de retrotransposons LINE e LTR (CHALOPIN *et al.*, 2015; HELLSTEN *et al.*, 2010).

Em *N. parkeri* a predominância é de retrotransposons, em sua maioria LTRs (SUN *et al.*, 2015). Para Sun e colaboradores (2015), talvez esse seja o motivo do genoma de *N. parkeri* ser maior em relação ao de *X. tropicalis*, devido ao mecanismo de transposição desses elementos (copia e cola), aumentando assim o número de cópias e consequentemente o tamanho do genoma.

Em *X. laevis*, especialmente, foram identificados elementos pertencentes a superfamília *SINE* e elementos do tipo MITEs. Esses elementos foram achados associados a 10 genes e em regiões regulatória 5' e 3', sugerindo que eles têm função na regulação dos genes (LEPETIT *et al.*, 2000). Outros elementos também descritos no genoma dessa espécie foram o *Tx1* e o *Tx2* (Classe I, ordem LINE). Essas cópias mostram especificidade no sítio de inserção por pentanucleotídeo TTTAA e possui TIRs de 19 pares de bases (pb) e TSDs de 4 pb. Elementos similares também são encontrados em *X. borealis* e *X. mulleri*, mostrando que a diversificação desse elemento pode ter ocorrido no ancestral em comum dessas espécies (GARRETT; CARROLL, 1986; GARRETT; KNUTZON; CARROLL, 1989).

Além desses, também foi descrito no genoma de *X. laevis* o transposon de DNA *Xstir*, um elemento do tipo MITE, que estima-se que existam aproximadamente um milhão de cópias por genoma ocupando 2% do conteúdo total desse genoma, um valor alto quando comparado com a quantidade de elementos transponíveis em outros indivíduos, como mamífero, por exemplo, onde prevalece retrotransposons (HIKOSAKA; YOKOUCHI; KAWAHARA, 2000). Esse elemento também foi encontrado em *X. borealis* e *X. tropicalis*, mas não foi encontrado na ordem Urodela e nem em mamíferos. Os fragmentos obtidos por clonagem genômica do elemento são indistinguíveis entre *X. borealis* e *X. laevis*, enquanto em *X. tropicalis* mostra diferenças entre os tamanhos das repetições e na identidade da sequência, sugerindo que um evento extenso de *bursts* de *Xstir* ocorreu em um ancestral comum de *X. borealis* e *X. laevis*, espécies que tiveram sua divergência pelo menos 15-20 milhões de anos atrás, enquanto em *X. tropicalis* aconteceu outro evento de amplificação, depois da separação do grupo, que ocorreu pelo menos 110-120 milhões de anos atrás (HIKOSAKA; YOKOUCHI; KAWAHARA, 2000). Essa conservação de elementos transponíveis do tipo MITE, sugere que sua atividade está sendo conservada pela seleção natural, consequentemente sendo vantajosa para o hospedeiro. Um modo de burlar a seleção é através da domesticação, onde os genes que compõem o elemento são descaracterizados e geralmente perdem sua capacidade de catalisar a transposição, sendo co-optados para uma

nova função no genoma. Esse exemplo pode ser visto no elemento Xmim, um membro da família T2 do tipo MITE, Superfamília Kolobok, presente nas espécies de *X. tropicalis* e *X. laevis*, eles são os mais prevalentes e provavelmente se originaram antes da divergência entre as linhagens (HIKOSAKA; KONISHI, 2018).

Session *et al* (2016), na descrição da sequência genômica de *X. laevis*, utilizou a presença e similaridade cópias de elementos de transposição como argumento adicional na teoria de origem alotetraploide dessa espécie. A hipótese mais aceita é a de que a espécie *X. laevis* teria sido originada pela hibridação entre dois progenitores diploides de $2n=18$ cromossomos, denominados S e L, seguido de uma duplicação genômica ocorrida para balancear o pareamento meiótico, gerando um organismo $2n=36$ cromossomos. Segundo Session *et al* (2016), essa hipótese pode ser corroborada através da identificação de elementos transponíveis específicos para cada lote cromossômico ancestral, provenientes de cada um dos genomas dos progenitores diploides. *MITEs* derivados da superfamília *PIF/Harbinger* foram identificado nos cromossomos provenientes do progenitor ancestral L (denominado XI-TpL_harb) e S (XI-TpS_harb). Além desses elementos, o *Tc1/Mariner* também foi identificado exclusivamente no cromossomo proveniente do progenitor S (XI-TpS_mar), reforçando a hipótese de origem alotetraploide em *X. laevis* (SESSION *et al.*, 2016).

Tratando-se de espécies de anfíbios sul-americanos, apenas um trabalho de isolamento, caracterização e mapeamento cromossômico foi realizado até o momento (NASCIMENTO; BALDO; LOURENÇO, 2015). Nesse trabalho, os autores analisaram espécies do gênero *Physalaemus*, um grupo atrativo para estudos citogenéticos, porque muitos indivíduos cariotipados até o momento, de diferentes populações e linhagens, possuem o mesmo número diploide, com variações em seu padrão de distribuição de heterocromatina e NOR (região organizadora do nucléolo) (QUINDERÉ *et al.*, 2009). Foi verificada a presença do elemento Rex1 (LINE) clusterizado na região heterocromática do braço curto do cromossomo 3 de *Physalaemus ephippifer*, evidenciando o acúmulo dessas sequências em regiões de heterocromatina associado com sequências rDNA 5S (Nascimento *et al.*, 2015). Esse panorama de poucos trabalhos de TEs sugerem que uma investigação mais detalhada sobre evolução e a dinâmica desses elementos no genoma de anuros devem ser feitas.

TEs da ordem LINE

Na classe I são encontrados os elementos da ordem LINEs (*Long Interspersed Element*) (WICKER *et al.*, 2007), onde estão presentes diversas superfamílias que codificam uma ORF, com pelo menos uma RT e uma EN. Além disso, em algumas superfamílias (Jockey, L1 e I), pode ser encontrada uma segunda ORF similar ao gene *gag* de retrovírus. Outra característica importante é o promotor interno e as repetições *in tandem*, que podem ser substituídas por uma cauda poli (A), na extremidade 3' (KAPITONOV; TEMPEL; JURKA, 2009; JURKA *et al.*, 2007; LÓPEZ-FLORES; GARRIDO-RAMOS, 2012). A divisão mais completa e atual dos elementos do tipo LINE inclui 28 clados de 6 grandes grupos: CRE, R4, Hero, NeSL e R2 do grupo R2; RandI do grupo RandI; Proto1, L1, e Tx1 do grupo L1; RTETP, Proto2, RTE e RTE do grupo RTE; Oucast, Ingi, I, Nimb, TadI, Loa e R1 do grupo I; Jockey, Rex1, CR1, L2, L2A, L2B, Daphne e Crack do grupo Jockey (KAPITONOV; TEMPEL; JURKA, 2009).

O mecanismo de retrotransposição dos LINEs e inserção no genoma é chamado de *target-primed reverse transcription* (TPRT) e a interrupção prematura dessa transcrição reversa pode explicar porque a maioria dos LINEs possuem extremidade 5' truncadas (JURKA *et al.*, 2007; LÓPEZ-FLORES; GARRIDO-RAMOS, 2012).

Esses elementos são menos comuns em plantas, mas abundantes nos genomas de aves (6, 4% no genoma de galo) e mamíferos, representando 18% do genoma de cães, 17% no genoma de humanos e 29% do genoma de gambás (LÓPEZ-FLORES; GARRIDO-RAMOS, 2012). No genoma de *X. tropicalis*, os LINEs são os retroelementos mais abundantes, ocupando 20% do genoma (CHALOPIN *et al.*, 2015; HELLSTEN *et al.*, 2010; SUN *et al.*, 2015), e no genoma de *N. parkeri* os LINEs estão em segundo em abundância, perdendo apenas para os retrotransposon LTRs (SUN *et al.*, 2015).

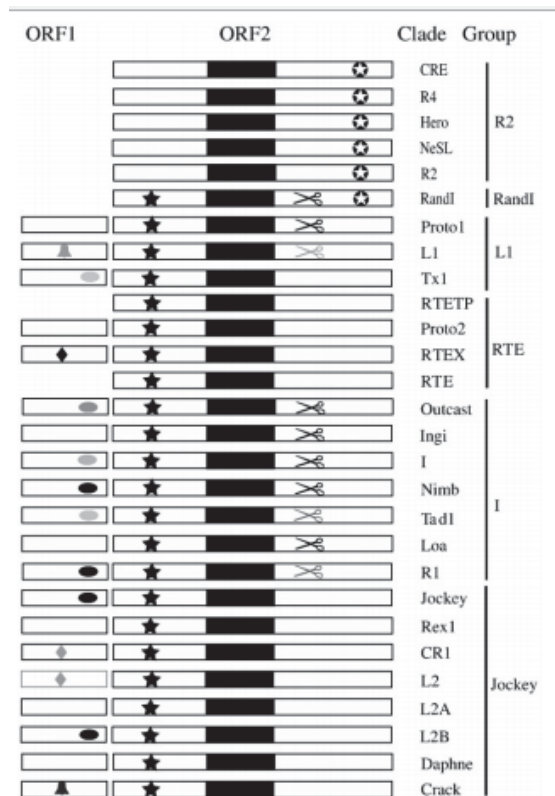


Figura 1. Diversidade molecular dos elementos tipo LINE, evidenciando suas superfamílias e as diferenças entre as estruturas moleculares, como presença de duas ORFs representadas pelos retângulos brancos. Na ORF 2 os retângulos pintados de preto representam a RT, as estrelas representam os domínios das diferentes endonucleases, as tesouras representam a RNaseH. Na ORF1 os desenhos de sinos e diamantes simbolizam os domínios das esterases, L1-like e RRM e as figuras ovais representam o domínio dedo de zinco. Note que a ORF1 só representada para alguns clados, pois não está presente em todos. Figura retirada de KAPITONOV; TEMPEL; JURKA, 2009.

Um elemento especialmente interessante do tipo LINE, são os genericamente classificados como Rex, que são amplamente distribuídos e estudados em peixes teleósteos, possuindo região codificadora para EN e RT (VOLFF *et al.*, 1999, 2001; VOLFF; KORTING; SCHARTL, 2000). Esses elementos foram primeiramente identificados em *Xiphophorus maculatus*, sendo nomeados de acordo com a ordem em que foram descobertos, como Rex1 (Retroelemento de *Xiphophorus* 1), Rex3 (Retroelemento de *Xiphophorus* 3), Rex5 (Retroelemento de *Xiphophorus* 5) e Rex6 (Retroelemento de *Xiphophorus* 6) (VOLFF *et al.*, 1999, 2001; VOLFF; KORTING; SCHARTL, 2000). Apesar de esses elementos compartilharem uma denominação semelhante, nenhum trabalho tem indicado que se tratem de um mesmo grupo monofilético.

Grupo de estudo: a família Pipidae

A família Pipidae, da ordem Anura, reúne atualmente quatro gêneros: *Hymenochirus* (Boulenger, 1896), *Pipa* (Laurenti, 1768), *Pseudohymenochirus* (Chabanaud, 1920) e *Xenopus* (Wagler, 1827) (FROST, 2018). Alguns autores ainda dividem o gênero *Xenopus* em dois, utilizando como critério análise filogenética e o número diploide ancestral, reconhecendo as espécies portadoras $2n=20$ (e seus múltiplos, tratando-se de espécies poliploides) sob a denominação *Silurana* enquanto espécies derivadas de um ancestral de $2n=18$, e seus múltiplos, sob o a denominação *Xenopus*.

As reconstruções filogenéticas mais robustas apontam que o gênero *Pipa*, que é o foco desse trabalho, é grupo-irmão dos demais pipídeos africanos (IRISARRI *et al.*, 2011). O gênero *Pipa* é composto por 7 espécies (FROST, 2018): *P. arrabali* Izecksohn, 1976, *P. aspera* Mueller, 1924, *P. myersi* Trueb, 1984, *P. parva* Ruthven & Gaige, 1923, *P. snethlageae* Muller, 1914, *P. pipa* (Linnaeus, 1758) e *P. carvalhoi* (Miranda-Ribeiro, 1937), distribuídas na América Central e do Sul (CANNATELLA; TRUEB, 1988).

Esse trabalho tem como objetivo de estudo o genoma da espécie *P. carvalhoi*, que foi recentemente sequenciado pelo nosso grupo de pesquisa (Laboratório de Citogenética Evolutiva e Conservação Animal, Universidade Federal do Paraná). Ela é a única espécie do gênero que ocorre em domínios não-Amazônicos, com relatos de sua presença tanto nos biomas de Caatinga e Mata Atlântica, que ocorre do estado do Ceará até o Espírito Santo. Essa espécie habita principalmente córregos e pântanos, sendo também conhecida por ocupar áreas que sofreram impacto ambiental, como lagoas artificiais (FROST, 2018; SILVA *et al.*, 2010).

2. OBJETIVOS

2.1. Objetivo geral

Analisar as relações evolutivas de elementos classificados como Rex da ordem LINE, atestando se realmente esses elementos estão presentes no genoma de *Pipa carvalhoi* e realizar o mapeamento cromossômico das cópias desses elementos em *P. carvalhoi* e *X. tropicalis*.

2.2. *Objetivos específicos*

- Buscas *in silico* dos retroelementos da ordem LINE denominados Rex no genoma de *P. carvalhoi*,
- Caracterização das sequências dos elementos da ordem LINE identificados e comparação desses elementos com os de *Xenopus tropicalis*.
- Análise evolutiva dos elementos LINE recuperados do genoma de *P. carvalhoi* juntamente com elementos LINE de diferentes superfamílias.
- Mapeamento cromossômico por hibridação *in situ* fluorescente (FISH) de sequências encontradas e selecionadas na análise *in silico* nos cariótipos de *P. carvalhoi* e em *X. tropicalis* e discussão sobre a localização cromossômica nessas espécies

CAPÍTULO I

Estudo dos retroelementos da ordem LINE no genoma de *Pipa carvalhoi* (Pipidae: Anura), com ênfase em elementos Rex

Joana de Moura Gama¹, Adriana Ludwig², Camilla Borges Gazolla¹, Shirlei Maria Recco-Pimentel³ & Daniel Pacheco Bruschi¹.

1- Departamento de Genética, Universidade Federal do Paraná (UFPR); 2- Instituto Carlos Chagas; 3- Departamento de Biologia Estrutural e Funcional, Universidade Estadual de Campinas (UNICAMP). Autor correspondente: adriludwig@gmail.com e danielpachecobruschi@gmail.com

Resumo: Elementos de transposição são sequências repetitivas que perfazem grande parte do genoma de eucariotos, sendo considerados importantes no papel da evolução e funcionamento dos genomas. A classificação dessas sequências ainda é um desafio devido à diversidade de elementos existente. Os retrotransposons perfazem grande parte dos genomas de eucariotos, incluindo os retroelementos do tipo LINE, objetos desse estudo. Uma maneira de classificá-los é através da análise de sequências da Transcriptase reversa (RT), o que foi feito nesse trabalho utilizando genomas de anuros, com enfoque na espécie *Pipa carvalhoi*, uma espécie endêmica da Mata atlântica pertencente à família Pipidae. Foram investigados elementos genericamente nomeados Rex (Rex1, Rex3, Rex5 e Rex 6), por meio de BLAST no genoma dos anuros e as cópias encontradas foram caracterizadas para a presença de fases abertas de leitura e regiões codificadoras de domínios proteicos conhecidos, como domínios. Os resultados corroboram que os elementos Rex pertencem a clados/superfamílias distintas (Rex1 – clado Rex1, Rex3 – clado RTE, Rex5 – clado L2 e Rex6 – clado R4). De forma geral, foi possível ter um panorama da presença desses elementos nas espécies de anuros estudados, sendo que eles parecem ser muito antigos nos genomas de vertebrados. Um possível evento de transferência horizontal de um elemento do clado RTE entre *P. carvalhoi* e *Helobdella robusta* (hirudínea) foi testado, porém não corroborado. Eventos mais recentes de retrotransposição de Rex1 e L2 parecem ter ocorrido em *P. carvalhoi* e vários eventos de perda de sequências devem ter acontecido em algumas linhagens de anuros para alguns elementos. Esse panorama da distribuição nos mostra que os TEs Rex não formam um grupo monofilético, além disso, a evolução desses elementos pode ajudar no entendimento do papel dessas sequências na diversificação dessas espécies.

Palavras-chaves: LINE, Anura, Pipidae.

INTRODUÇÃO

Elementos genéticos móveis correspondem a uma parcela significativa do genoma de organismos eucariotos e são considerados importantes substratos evolutivos na diversificação dos genomas (BIÉMONT; VIEIRA, 2006; FARRÉ *et al.*, 2011; GRAY, 2000).

Apesar de ainda restritos a poucas espécies, a análise da fração repetitiva do genoma de anfíbios anuros não foge à regra dos demais vertebrados (CHALOPIN *et al.*, 2015; FESCHOTTE; PRITHAM, 2007; KAMINKER *et al.*, 2002; KEINATH *et al.*, 2015; SOTERO-CAIO *et al.*, 2017), com massiva invasão e expressiva diversidade desses elementos em seus genomas (CHALOPIN *et al.*, 2015; HELLSTEN *et al.*, 2010; SUN *et al.*, 2015).

A invasão desses elementos pode ocorrer em um genoma específico e os mesmos são transferidos verticalmente às linhagens derivadas. Transferência Horizontal (TH) também é um mecanismo responsável pela disseminação dessas sequências entre os genomas, podendo ocorrer a transferência desses elementos entre genoma de diferentes espécies, normalmente aquelas que ocupam o mesmo nicho, porém filogeneticamente pouco relacionados (SILVA; MARGARET, 2000; WALLAU; VIEIRA; LORETO, 2018) .

Retrotransposons (ditos elementos de Classe I) perfazem grande parcela dos elementos de transposição (TEs) da maioria dos genomas (CHALOPIN *et al.*, 2015; SUN *et al.*, 2015) e devido ao seu mecanismo de transposição, são apontados como os principais responsáveis pelo aumento do tamanho de alguns genomas (SUN *et al.*, 2015). Dentre eles, a ordem LINE (WICKER *et al.*, 2007) tem destaque principalmente por estar presente em praticamente todos os organismos eucariontes e com grande diversidade de famílias (CHALOPIN *et al.*, 2015; HELLSTEN *et al.*, 2010).

Os elementos do tipo LINE são caracterizados pela presença de pelo menos uma ORF (Pol-like), codificando domínios de transcriptase reversa (RT) e endonuclease (EN). Alguns tipos de LINE também possuem uma ORF adicional *upstream* denominada gag-like (JURKA *et al.*, 2007; LÓPEZ-FLORES; GARRIDO-RAMOS, 2012). Além disso, é possível encontrar uma extremidade 3' com repetições *in tandem* e regiões ricas em Adenosina (A) (VOLFF *et al.*, 1999, 2001; VOLFF; KORTING; SCHARTL, 2000; WICKER *et al.*, 2007).

Segundo a classificação proposta por Wicker *et al.* (2007), a ordem LINE reúne cinco superfamílias (R2, L1, RTE, I e Jockey). De acordo com essa classificação, as superfamílias são divididas por sua estratégia de replicação, estrutura das proteínas e domínios, e característica das *Target Site Duplications* (TSDs). Dentro de cada superfamília, os elementos são separados em famílias de acordo como grau de conservação das sequências de DNA com mais de 80 pares de bases e no mínimo de 80% de similaridade em pelo menos 80% da sua região codificante ou nos domínios internos, ou nas suas regiões terminais (WICKER *et al.*, 2007). Outros autores dividem os elementos do tipo LINE em um maior

número de clados/superfamílias, como CR1, CRE, I, Jockey, L1, NeSL, R2, R4, RandI, Rex1, RTE e Tx1 (KAPITONOV; JURKA, 2008).

Com o aumento do número de genomas sequenciados e a disponibilização dos mesmos, novos TEs têm sido descritos e classificados, usando principalmente, como base a proposta de classificação de Wicker *et al* (2007) ou Kapitonov & Jurka (2008). Uma maneira para classificação dos LINEs é através da análise de similaridade da Transcriptase reversa (RT), o domínio mais utilizado para classificações de retrotransposons, pois é o único compartilhado entre todos esses retroelementos e tem sido muito útil nessa abordagem, que demonstra ser um desafio (KAPITONOV; JURKA, 2008; MALIK; BURKE; EICKBUSH, 1999). A divisão mais completa e atual dos elementos do tipo LINE inclui 28 clados de 6 grandes grupos: CRE, R4, Hero, NeSL e R2 do grupo R2; RandI do grupo RandI; Proto1, L1, e Tx1 do grupo L1; RTETP, Proto2, RTE e RTE do grupo RTE; Oucast, Ingi, I, Nimb, TadI, Loa e R1 do grupo I; Jockey, Rex1, CR1, L2, L2A, L2B, Daphne e Crack do grupo Jockey (KAPITONOV; TEMPEL; JURKA, 2009). Ainda, existe o clado Odin estabelecido por Volff *et al* (2004).

Elementos do tipo LINE são genericamente classificados como Rex, que são amplamente distribuídos em peixes teleósteos, possuindo região codificadora para EN e RT (VOLFF *et al.*, 1999, 2001; VOLFF; KORTING; SCHARTL, 2000). Esses elementos foram primeiramente identificados em *Xiphophorus maculatus*, sendo nomeados de acordo com a ordem em que foram descobertos, como Rex1 (Retroelemento de *Xiphophorus* 1), Rex3 (Retroelemento de *Xiphophorus* 3), Rex5 (Retroelemento de *Xiphophorus* 5) e Rex6 (Retroelemento de *Xiphophorus* 6) (VOLFF *et al.*, 1999, 2001; VOLFF; KORTING; SCHARTL, 2000). Esses elementos são eventualmente classificados como todos de uma mesmo grupo hipotético, muitas vezes denominado como superfamília Rex (CARDUCCI *et al.*, 2018; FAVARATO *et al.*, 2017), um erro comum devido ao seu nome e ao complexo sistema de classificação dos TEs. No entanto, não há evidência de um agrupamento natural dessas sequências, sendo que Rex1 pertence ao clado Rex1 (VOLFF; KORTING; SCHARTL, 2000), Rex3 pertence ao clado RTE (LOVŠIN; GUBENŠEK; KORDIŠ, 2001), Rex5 pertence ao clado L2 (LOVŠIN; GUBENŠEK; KORDIŠ, 2001) e Rex6 pertence ao clado R4 (BURKE *et al.*, 2002). Embora inicialmente considerados elementos específicos de peixes teleósteos (CARDUCCI *et al.*, 2018; VOLFF *et al.*, 1999; VOLFF; KORTING; SCHARTL, 2000), sequências homólogas têm sido encontradas também em outros organismos, incluindo anfíbios (NASCIMENTO; BALDO; LOURENÇO, 2015).

Em relação ao genoma de anfíbios é possível observar grande diversidade de elementos da ordem LINE nas espécies *Xenopus tropicalis*, *Rhinella marina* e *Nanorana parkeri* (EDWARDS *et al.*, 2018; HELLSTEN *et al.*, 2010; SUN *et al.*, 2015). Nesse trabalho, investigamos a diversidade de elementos de transposição LINE no genoma de Anuros já sequenciados até o momento, com foco em sequências homólogas aos LINEs denominados Rex presentes no genoma “rascunho” de *Pipa carvalhoi* (Anura), uma espécie de Pipídeo Neotropical distribuída pelos biomas da Caatinga e Mata Atlântica, desde o estado do Ceará até o Espírito Santo (FROST, 2019). Essa espécie, assim como os demais da família Pipidae (como representantes do gênero *Xenopus*), possui hábitos aquáticos, habitando córregos e pântanos, sendo também conhecida por ocupar áreas de impacto ambiental, como lagoas artificiais (FROST, 2019; SILVA *et al.*, 2010).

MATERIAL E MÉTODOS

Busca de TEs no genoma de P. carvalhoi

Utilizamos a versão rascunho do genoma de *P. carvalhoi*, recentemente sequenciado pelo Laboratório de Citogenética Evolutiva e Conservação Animal da Universidade Federal do Paraná, Brasil. O genoma sequenciado na plataforma *Illumina HiSeq-400*, 46x de cobertura, foi montado utilizando o algoritmo *Velvet* e está em nível de *contigs* (2.4 Gpb, 5.029.551 *contigs*). Devido à grande fragmentação do genoma, *contigs* menores que 400 pb foram removidos da análise, resultando em um *dataset* de 1.3 Gpb. A análise do conteúdo de TEs total dos genomas foi realizada utilizando o pipeline PiRATE (*Pipeline to Retrieve and Annotate Transposable Elements*), disponível para download em <http://www.seanoe.org/data/00406/51795/> (BERTHELIER *et al.*, 2018). As buscas no PiRATE seguiram as seguintes etapas: (1) identificação de sequências repetitivas por diferentes ferramentas com base em similaridade, estrutura e repetitividade, (RepeatMasker, TE-HMMER, LTRharvest, Helsearch, MGEScan-nonLTR, MITE-Hunter, SINE-Finder, TEdenovo e Repeat Scout); (2) concatenação das bibliotecas de sequências identificadas por cada programa e filtragem pela ferramenta CD-HIT-est para eliminação de redundância; e (3) classificação automática das sequências pela ferramenta PASTEC, que é baseada na classificação de Wicker *et al.* (2007).

Busca das cópias dos retroelementos no genoma dos Anuros

Foram recuperados os retroelementos genericamente denominados Rex (Rex1, 3, 5 e 6), realizamos buscas utilizando a ferramenta tBLASTn (ALTSCHUL *et al.*, 1997) no genoma de *P. carvalhoi* localmente. Para essas buscas, utilizamos como *query* todos os elementos identificados como Rex depositados no banco de dados do Repbase (JURKA *et al.*, 2005), resultando em um conjunto de aproximadamente 80 *queries* (Rex1, Rex3, Rex5 e Rex6). Baseados nos resultados das buscas por tBLASTn escrevemos código em linguagem Python para recuperar os contigs do genoma de *P. carvalhoi*. Para cada *query*, foram analisados os 15 *contigs* com maior significância (maior score e *e-value* menor que 10^{-4}). A presença de ORFs foi analisada pela ferramenta do NCBI, ORFinder (<https://www.ncbi.nlm.nih.gov/orffinder/>), e os domínios proteicos foram preditos através da ferramenta CD-Search do NCBI (MARCHLER-BAUER; BRYANT, 2004). Os elementos foram considerados completos quando possuíam uma ORF com domínios de uma EN e RT que não estavam truncados ou degenerados.

Para simplificar a busca dos elementos nos demais genomas de anuros e evitar redundância nas buscas, realizamos uma seleção de sequências para serem usadas como *queries*. Uma árvore filogenética preliminar, baseada na RT, foi gerada por Análise Bayesiana (BA) no software MrBayes 3.2.6 (RONQUIST *et al.*, 2012), sob um modelo misto, realizando 10,000,000 de gerações de cadeia de Markov Monte Carlo (MCMC) e amostrando árvores a cada 1,000 gerações. Os 25% dos resultados iniciais (*burn-in*) foram descartados (Anexo I). Nessa análise foram incluídas as sequências obtidas de *P. carvalhoi* juntamente com as *queries* e sequências conhecidas de diferentes tipos de LINES. O alinhamento múltiplo dos aminoácidos foi realizado utilizando a plataforma MAFFT versão 7 (<https://mafft.cbrc.jp/alignment/software/>), utilizando a opção “ L-INS-i (recomendado para <200 sequências com um domínio conservado e longos *Gaps* (KATO; ROZEWICKI; YAMADA, 2017). A partir dessa árvore, foram escolhidas as *queries* provenientes do genoma de *P. carvalhoi* presentes em clados distintos. Essas sequências ainda foram alinhadas no MUSCLE *online* (<https://www.ebi.ac.uk/Tools/msa/muscle/>) (MADEIRA *et al.*, 2019) e a divergência entre as cópias foi calculada pelo programa GeneDoc 2.7 (NICHOLAS; NICHOLAS; DEERFIELD, 1997). As sequências com similaridade menor que 60% entre si foram mantidas, sendo que apenas o clado L1 forneceu duas *queries* (L1A e L1B). A busca no genoma dos demais representantes de anuros, disponíveis no GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) foi feita através de tBLASTn local e *online* (descrição adicional sobre esses genomas estão compiladas na Tabela 1), e as regiões

contendo os elementos foram recuperadas pela ferramenta *online* ou com auxílio de script em Pearl.

Tabela 1: Informações dos genomas de Anuros disponíveis que foram utilizados nesse trabalho.

Família	Espécie	N M	Co b	Nº. Sc	Nº. Co	Sequenciamento	Número de Acesso GenBank
Pipidae	<i>Xenopus tropicalis (Xt)*</i>	Cr	111x	167	851	PacBio Sequel e Illumina Hi Seq	GCA_000004195.4
	<i>Xenopus laevis (Xl)*</i>	Cr	30x	108,033	324,061	Illumina HiSeq 2000	GCA_001663975.1
Megophryidae	<i>Leptobrachium leishanense (Lle)*</i>	Cr	80x	5,302	8,584	PacBio RSII	GCA_009667805.1
Pyxicephalidae	<i>Pyxicephalus adspersus (PyA)*</i>	Cr	189x	5,411	117,078	Illumina HiSeq X	GCA_004786255.1
Bufonidae	<i>Rhinella marina (Rm)*</i>	Co	141x	-	31,391	sem informação	GCA_900303285.1
Dicroglossidae	<i>Nanorana parkeri (Np)*</i>	Sc	83x	25,188	147,513	Illumina Genome Analyzer II	GCA_000935625.1
Scaphiopodidae	<i>Spea multiplicata (Sm)*</i>	Sc	21x	49,736	74,867	PacBio Sequel	GCA_009364415.1
	<i>Scaphiopus holbrookii (Sh)**</i>	Sc	39x	1,509,266	1,509,291	Illumina NextSeq	GCA_009364455.1
Ranidae	<i>Rana catesbeiana (Lithobates catesbeianus) - (Lca)*</i>	Sc	66x	1,544,635	2,124,505	Illumina HiSeq	GCA_002284835.2
Dendrobatidae	<i>Oophaga pumilio (Op)**</i>	Sc	136x	125,383	637,067	Illumina HiSeq	GCA_009801035.1

Nota: Na tabela evidenciamos a família, o nome da espécie e código referente a ela entre parênteses. * - Blast online; ** - Blast local; NM – nível de montagem; Cr – cromossomo; Co – *contig*; Sc – *scaffold*; Cob - cobertura do sequenciamento; Nº. S - número de *scaffolds*; Nº. C - número de *contigs*.

Foram analisados os 10 *hits* com maior *score* e *e-value* significativos (menor que 10^{-4}) com intuito de recuperarmos a sequência dos domínios da RT. Para isso, a presença de ORFs foi analisada no ORFinder e o domínios preditos pela ferramenta CD-Search. As sequências dos Anuros, exceção de *P. carvalhoi*, foram nomeadas seguindo uma ordem: nome do elemento, código da espécie (Tabela 1), número da cópia. Como exemplo REX1Xtc1 o elemento REX1 de *X. tropicalis*, cópia 1.

Análise das sequências LINES encontradas em anuros

As inferências de relacionamento evolutivo entre as cópias de elementos da ordem LINE recuperados dos genomas dos anuros foram conduzidas utilizando a sequência de aminoácidos do domínio completo da RT. Completando a matriz, foram incluídas sequências de elementos recuperados a partir do banco de dados do Repbase pertencendo a diferentes grupos de LINES já estabelecidos. O alinhamento múltiplo dos aminoácidos foi realizado utilizando a plataforma MAFFT versão 7. O alinhamento foi visualizado no GeneDoc 2.7 e inspecionado manualmente. A ferramenta trimAl disponível *online* (<https://ngphylogeny.fr>), foi utilizada para isolar regiões de alinhamentos mais confiáveis com os seguinte parâmetros: *GAP threshold* 0,9, *Similary thresold* 0,0 e *Consistency threshold* 3,0, (CAPELLA-GUTIÉRREZ; SILLA-MARTÍNEZ; GABALDÓN, 2009).

Uma árvore preliminar por *Neighbour Joining* (NJ) utilizando modelo de *Poisson* (Anexo V), com 1000 réplicas de *bootstrap* foi obtida no programa MEGA X (KUMAR *et al.*, 2018). Esta árvore foi utilizada com intuito de filtrar sequências muito próximas que agrupavam no mesmo clado, por serem possivelmente duas cópias de um mesmo elemento, dessa maneira simplificamos a matriz de dados. As sequências de *P. carvalhoi* foram todas mantidas, independente da similaridade. Após essa filtragem, realizamos a busca pelo modelo evolutivo mais adequado aos dados, utilizando programa Mega X. A relação entre as sequências de interesse foi obtida (Figura 1) através do método de máxima verossimilhança (ML) pela ferramenta RAxML-HPC BlackBox (STAMATAKIS, 2014), implementada no portal CIPRES (MILLER; PFEIFFER; SCHWARTZ, 2010). Análise Bayesiana também foi utilizada para inferência das relações evolutivas entre as sequências, sendo implementada no MrBayes 3.2.6 com os mesmos parâmetros já citados no tópico anterior (Anexo II). As árvores foram visualizadas e editadas no ITOL (*Interactive Tree of life*- <https://itol.embl.de/>) (LETUNIC; BORK, 2019) e enraizada pelo *cluster* CRE, como proposto por Malik *et al.* (1999) e Eickbush e Malik (2002), sendo que os *clusters* também foram nomeados de acordo

com esses trabalhos.

Baseados na lista dos genomas incluídos no presente estudo, utilizamos a plataforma Timetree (<http://www.timetree.org/book>) (HEDGES *et al.*, 2015) para a construção de um cladograma consenso capaz de refletir o relacionamento filogenético entre as espécies incluídas no estudo. O cladograma gerado por tal plataforma é resultado da compilação de dados de árvores filogenéticas publicada, resultando em uma árvore consenso inferida para os táxons incluídos na análise. Devido à reduzida amostragem de espécies com genoma sequenciado frente a real diversidade de anuros conhecida, nosso intuito com esse cladograma foi recuperar uma árvore que refletisse o posicionamento mais próximo dessas espécies na árvore da vida de anuros e assim mapear dados de presença/ausência de superfamílias/clados de TEs no genoma dessas espécies (Figura 2).

Teste de Transferência Horizontal (TH)

Um possível caso de transferência horizontal entre o elemento RTE de *Helobdella robusta* e *P. carvalhoi* foi testado por uma abordagem amplamente utilizada, pelo fato de apresentarem uma incongruência filogenética porque agruparem em um mesmo *cluster* de sequências (BARTOLOMÉ; BELLO; MASIDE, 2009; LUDWIG; VALENTE; LORETO, 2008; SILVA; MARGARET, 2000), comparando a taxa de substituição sinônima (dS) (medida de evolução neutra) entre o retroelemento e genes órtologos constitutivos das espécies envolvidas. Devido à ausência de anotação gênica em *P. carvalhoi*, genes de *X. tropicalis* foram usados, sendo que resultados similares de dS são esperados, pois o tempo evolutivo de separação entre *H. robusta* e *P. carvalhoi* é o mesmo que de *H. robusta* e *X. tropicalis*, embora diferentes taxas de substituição após a especiação que separou *Xenopus* e *Pipa* poderiam influenciar o resultado. Oito genes ortólogos de cópia única de *H. robusta* e *X. tropicalis* (Tabela 2) foram escolhidos aleatoriamente com ajuda da ferramenta OrthoDB (<https://www.orthodb.org/>). O alinhamento dos códons de cada gene foi feito pela ferramenta MUSCLE no programa MEGA X, após a tradução da sequência de nucleotídeos. A estimativa de dS de cada gene foi calculada utilizando o método Nei & Gojobori (1986) também no MEGA X. Levando em conta que o uso preferencial de códons tem efeito direto sobre o dS, estimamos o uso de códons pelo cálculo do ENC (número efetivo de códons) e CBI (índice de códon *bias*) na ferramenta DAMBE (*Data Analysis in Molecular Biology and Evolution*) (XIA, 2018).

Tabela 2: Genes ortólogos em *X. tropicalis* e *H. robusta* usados para testar a hipótese de TH.

Gene	Nome	Número de acesso <i>H. robusta</i>	Número de acesso <i>X. tropicalis</i>
Gene 1	<i>Armadillo repeat containing 3</i>	XM_009029350	BC067323
Gene 2	<i>Glutamine-dependent NAD (+) synthetase</i>	XM_009018684	BC161121
Gene 3	<i>Glutamate-cysteine ligase</i>	XM_009016541	BC121418
Gene 4	<i>Hypoxia up-regulated protein 1</i>	XM_009018713	BC121233
Gene 5	<i>Dual serine/threonine and tyrosine protein kinase</i>	XM_009012804	AY364233
Gene 6	<i>Rab-GTPase-TBC domain</i>	XM_009024082	BC136015
Gene 7	<i>rspryl</i>	XM_009012401	BC123024
Gene 8	<i>Protein SDA1 homolog</i>	XM_009011314	BC135943

Para verificar a presença de sequências homólogas ao elemento encontrado em *H. robusta* em espécies próximas, realizamos uma busca por Blastn contra o genoma de *Eisenia fetida*, uma espécie do mesmo Filo (Annelida) e Classe (Clitellata). A sequência mais similar teve o domínio RT isolado e incluído na matriz para as análises filogenéticas.

RESULTADOS E DISCUSSÃO

Draft do genoma de Pipa carvalhoi

Os reads obtidos a partir do sequenciamento do genoma de *P. carvalhoi*, feito através da plataforma Illumina, resultaram em uma montagem de 5.029.551 contigs e aproximadamente 2.4 Gpb. Apesar de o genoma estar com cobertura de 46X, devido ao nível de fragmentação, os contigs foram filtrados por tamanho, permanecendo apenas os fragmentos maiores de 400 pb (1,153,690 contigs, 1.3 Gpb de sequências). Desse modo, para a abordagem inicial dos estudos dos TEs utilizamos apenas uma parcela das sequências, porém, essa filtragem permitiu que as análises pudessem ser realizadas de forma mais rápida.

Por meio do pipeline PIRATE, conseguimos obter um panorama global do conteúdo de TEs e sequências repetitivas presentes na montagem disponível do genoma de *P. carvalhoi*. Os dados encontrados revelam uma grande diversidade de elementos de transposição, pertencentes a diferentes ordens. Foram identificados 67,847 fragmentos

correspondentes a sequências repetitivas após a filtragem para eliminação de redundância (ou seja, são fragmentos distintos de sequências repetitivas, não representando sua abundância). Destes 50,472 fragmentos são Retrotransposons e 10,577 são Transposons de DNA. Ainda, foram encontrados seis fragmentos de genes (Potential Host Gene), 4,145 de Microsatélites (SSR) e 2,667 fragmentos de sequências repetitivas não categorizadas (noCat), podendo ser DNA satélites ou minissatélites, famílias multigênicas ou até mesmo TEs degenerados. O futuro mapeamento desses fragmentos no genoma permitirá obter informações da sua cobertura, ou seja, qual a porcentagem do genoma de *P. carvalhoi* que essas sequências ocupam.

A tabela 3 mostra um resumo das diferentes superfamílias de TEs identificadas. Foram recuperadas cópias com similaridade com retrotransposons das ordens DIRS, PLE, LTR, SINE, LINE. Dentre as sequências com similaridade com elementos da ordem LTR foram identificadas sequências de quatro superfamílias: *Gypsy*, *Bel-Pao*, *ERV* e *Copia*.

Em relação aos LINES, nossa análise recuperou quatro superfamílias: *Jockey*, *L1*, *RTE* e *I*. Dentre os fragmentos classificadas como *Jockey*, podemos observar sequências com similaridade a elementos que pertencem a *clusters* distintos (CR1 e Rex1) e poderiam ser classificados em diferentes superfamílias conforme a classificação de Kapitonov e Jurka (2008). O mesmo acontece com os elementos classificados como *L1*, que incluem sequências com similaridade a elementos Tx1, que formam um clado distinto (JURKA *et al.*, 2005).

Tabela 3. Classe, ordem e famílias dos elementos de transposição identificados no genoma de *Pipa carvalhoi*, através da *pipeline* PIRATE. A classificação desses elementos é baseada em Wicker *et al* (2007).

Classe	Ordem	Superfamília
Classe I- Retrotransposons	<i>DIRS</i>	<i>DIRS</i>
		<i>Gypsy</i>
		<i>Bel-Pao</i>
		<i>ERV</i>
	<i>LINE</i>	<i>Copia</i>
		<i>Jockey</i>
		<i>L1</i>
		<i>RTE</i>
		<i>I</i>

	<i>PLE</i>	<i>Penelope</i>
Classe II- Transposon de DNA	<i>TIR</i>	<i>Tc1- Mariner</i> <i>PiggyBac</i> <i>hAT</i> <i>NovoSib</i> <i>PIF-Harbinger</i> <i>CACTA</i> <i>MuDR</i> <i>Sola</i> <i>IS3EU</i>
	<i>Maverick</i>	<i>Maverick</i>
	<i>Helitrons</i>	<i>Helitrons</i>
Total	7	22

Quando consideramos os elementos de classe II, nossos dados recuperaram as duas principais subclasses reconhecidas por Wicker *et al.* (2007), subclasse I (identificados elementos apenas da ordem TIR) e subclasse 2 (identificados elementos das ordens Helitron e Maverick). Dentre os TIRs foram identificados elementos com similaridade a sequências das superfamílias *Tc1-Mariner*, *PiggyBac*, *hAT*, *Novosib*, *PIF-Harbinger*, *CACTA*, *MuDR*, *Sola*, *IS3EU*.

Dentre elementos considerados não-autônomos, foram identificados 25 fragmentos de elementos MITEs (*miniature inverted-repeat transposable elements*) e 762 fragmentos de elementos TRIM (*terminal-repeat retrotransposons in miniature*), elementos de classe II e I, respectivamente.

De forma geral, identificamos mais de 20 superfamílias de TEs em *P. carvalhoi*, uma diversidade similar à encontrada em peixes teleósteos (24 superfamílias) e maior do que a encontrada em mamíferos (11-14 superfamílias), aves (7-9 superfamílias) e répteis (15-18 superfamílias) (CHALOPIN *et al.*, 2015).

A diversidade de TEs encontrada em *P. carvalhoi* é similar a encontrada em outros anuros como *N. parkeri* (23 superfamílias) (SUN *et al.*, 2015), e *X. tropicalis* (21 superfamílias) (CHALOPIN *et al.*, 2015). Comparando nossos dados a listagem de elementos identificados nos genomas de *N. parkeri* (SUN *et al.*, 2015) e *X. tropicalis* (HELLSTEN *et al.*, 2010) pudemos perceber a ausência cópias de elementos da Superfamília RTE pertencente a ordem LINE, no genoma de ambas as espécies, enquanto nossos dados identificaram a

presença de cópias desse elemento no genoma de *P. carvalhoi* (HELLSTEN *et al.*, 2010; SUN *et al.*, 2015).

Nossas análises identificaram um número baixo de fragmentos correspondentes a elementos SINE (49 fragmentos) e MITEs, elementos reconhecidamente de tamanho pequeno. Não podemos descartar a possibilidade de que uma diversidade maior desses elementos esteja presente em *P. carvalhoi* e tenha sido eliminada pela filtragem de sequências maiores que 400 pb em nossas análises.

Pelo fato da montagem do genoma de *P. carvalhoi* ainda não estar num nível adequado, juntamente com a filtragem de sequências mencionada e a ausência de um mapeamento dos TEs encontrados nesse genoma, torna impossível, nesse momento, comparar o conteúdo repetitivo do genoma de *P. carvalhoi* em termos de densidade.

O conteúdo de TEs tem sido utilizado como explicação para a diferença de tamanho entre os genomas, sendo o maior genoma aquele que possui maior porcentagem de retroelementos do que elementos de DNA, devido ao seu mecanismo de transposição popularmente conhecido como “copia-e-cola” (CHALOPIN *et al.*, 2015; SUN *et al.*, 2015). Em anuros essa estimativa foi proposta comparando os genomas de *N. parkeri* (2,5 Gb), onde temos prevalências de elementos de classe I, e *X. tropicalis* (1,5 Gb) com prevalência de elementos de classe II (SUN *et al.*, 2015). Desde essa comparação em 2015, tivemos o aumento no número de genomas de anuros sequenciados, porém, com carência de anotação das sequências repetitivas. Dessa maneira, podemos sugerir a importância de uma análise futura aprofundada do conteúdo de TEs e a sua relação com o tamanho do genoma de anuros.

Devemos levar em consideração que aqui trabalhamos com uma versão rascunho do genoma de *P. carvalhoi* e esses números de fragmentos tendem a mudar quando uma nova montagem for analisada, talvez até com a identificação de novas repetições. Essas análises serão refeitas em uma nova montagem do genoma que será obtida após sequenciamento pela plataforma *PacBio*.

Análise comparativa dos Elementos Rex

As buscas de sequências homólogas a elementos denominados Rex no genoma de *P. carvalhoi* recuperaram cópias com uma variedade em seu nível de integridade. Como esperado, várias queries relacionadas recuperaram os mesmos contigs/cópias. No total, foram

analisados 181 *contigs* e a grande maioria apresentou domínios para RT (sendo 44 desses o domínio completo e 91 incompletos ou truncados), apenas 18 desses *contigs* apresentaram domínio de RT e EN, sendo oito desses na mesma ORF e dez em ORFs separadas (Informação suplementar – Anexo III). Uma característica dos LINEs é a presença desses dois domínios em uma única ORF (JURKA *et al.*, 2007; MALIK; BURKE; EICKBUSH, 1999; VOLFF; KORTING; SCHARTL, 2000), indicando que várias dessas cópias identificadas em *P. carvalhoi* estão possivelmente degeneradas. Três *contigs* apresentaram apenas o domínio da EN e sete apresentaram outros domínios além da EN e RT, como: *PKc_like*, *Arginine methyltransferase-interacting protein*, *RING Zn-finger*, *chromosome segregation protein*, *zinc-binding in reverse transcriptase*, *universal minicircle sequence binding protein (UMSBP)*, *RecF/RecN/SMC N terminal domain*, todos degenerados (Anexo IV).

As sequências com domínio da RT completo ou com 80% da sequência completa (55 sequências) foram usadas para as análises filogenéticas. Para entender as relações entre essas sequências e suas respectivas *queries*, realizamos uma análise filogenética incluindo também sequências de elementos LINEs sabidamente pertencentes a diferentes clados/superfamílias e sequências recuperadas no genoma de anuros disponíveis no GenBank (Tabela 1). Um total de 537 novas sequências foram adicionadas a nossa matriz, juntamente com sequências provenientes do REPBASE (140 sequências), onde estavam representantes de elementos da ordem LINE em geral, totalizando uma matriz de 732 sequência, variando entre 300-500 aminoácidos, a qual foi utilizada para construção de uma árvore NJ (Anexo V). A partir dessa árvore as sequências de maior similaridade foram eliminadas, para melhor visualização e entendimento dos resultados, a nova árvore ML foi feita através de uma nova matriz com 329 sequências. Utilizamos os critérios empregados por Malik, *et al* (1999) e Eickbush e Malik, (2002) para determinar a classificação dos elementos em diferentes clados/grupos. Os *clusters* recuperados foram identificados com os nomes dos elementos que já estão melhores estudados e classificados (EICKBUSH; MALIK, 2002; MALIK; BURKE; EICKBUSH, 1999).

Apesar de grande parte das relações entre as superfamílias de TE não estarem bem suportadas e resolvidas (Figura 1), foi possível recuperar a origem comum com alto suporte para os principais *clusters* de LINEs. As cópias isoladas do genoma dos anuros foram agrupadas com elementos dos clados L1, TX1, RTE, Rex1, L2 e CR1. Os elementos pertencentes aos clados L1, TX1 e L2 estão presentes no genoma de todas espécies analisadas nesse trabalho (Figura 1). Já a arranjo obtido por BA (Anexo II) mostrou uma maior resolução

das relações entre os *clusters*. Apesar de clusterizados entre si, uma grande politomia reunindo elementos dos clados R1, Loa, Tad1, Jockey e Ingi foi observada em nossa árvore de sequências. Em sequência, dois outros grandes grupos se formam, um contendo NeSL, R2 e R4 monofiléticos, RTE, e Tx1 e L1 também monofiléticos, e outro contendo Odin, seguindo da ramificação de L2 e em sequência CR1 e Rex1 (Figura 1).

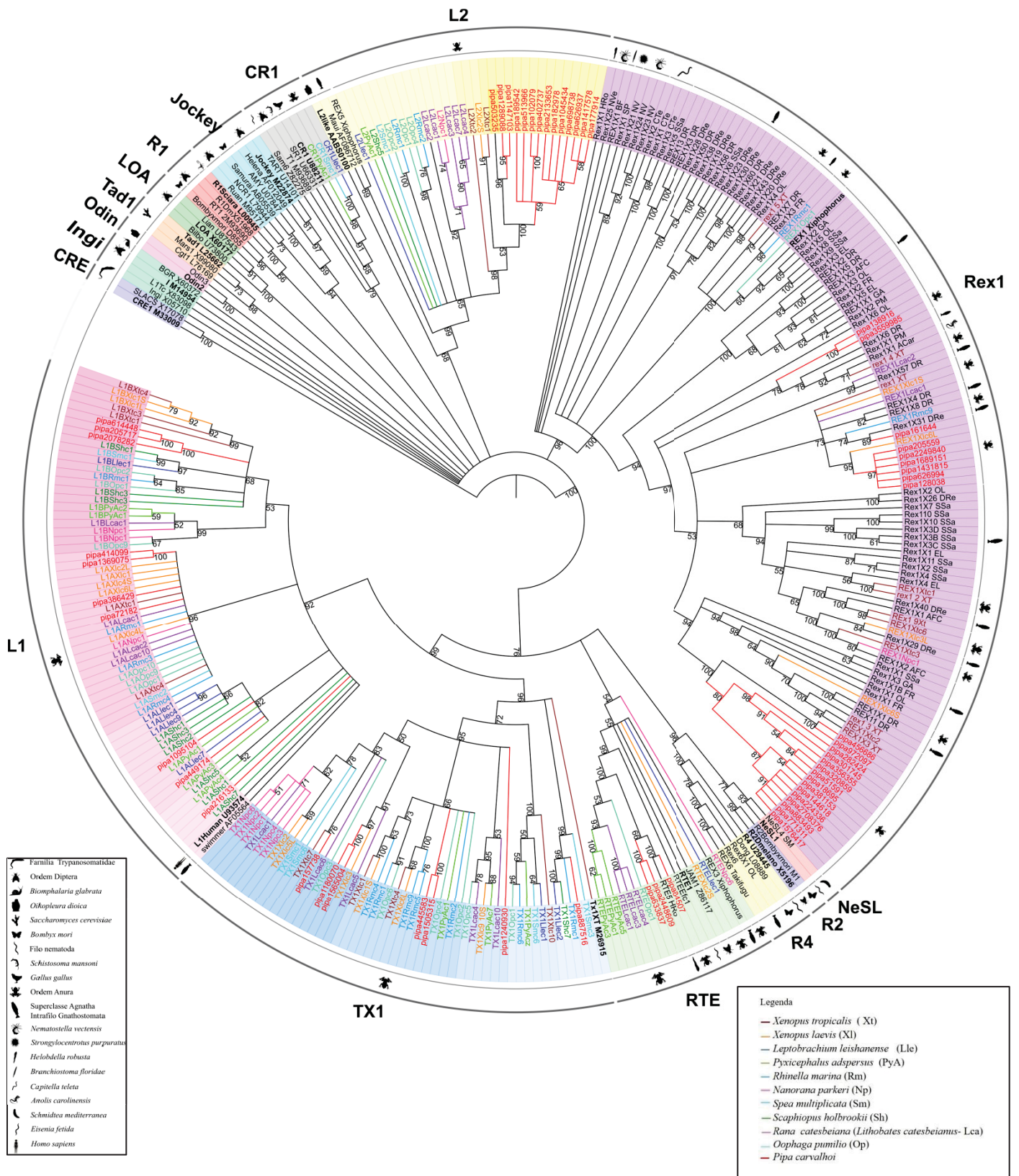


Figura 1- Árvore filogenética reconstruída por máxima verossimilhança (ML), modelo LG+G, a partir do domínio da RT dos elementos LINES recuperados no genoma de *P. carvalhoi*. Os valores de *bootstrap* são indicados próximo aos nós. Cada clado está representado por uma cor assim como cada espécie de anuro (conforme a legenda).

Em geral as sequências do genoma de *P. carvalhoi* agruparam com as *queries* utilizadas para recuperá-las, com algumas exceções. Todas as *queries* de Rex1 recuperaram sequências que agruparam no clado Rex1, essa condição também foi observada para a *query* de Rex5, onde todas as sequências agruparam no clado L2. Já as *queries* Rex3, Rex6 e Rex6-1_OL, recuperaram sequências que agruparam em mais de um clado (Tabela 4).

Tabela 4. *Queries* do Repbase utilizadas no tBLASTn contra o genoma de *P. carvalhoi* com os respectivos clados formados pelas sequências recuperadas desse genoma.

<i>Queries</i> do Repbase	Clados das sequências (seq.) recuperadas em <i>P. carvalhoi</i>
Rex1 (76 seq.)	Rex1 (22 seq.)
Rex3 (1 seq.)	L1 (3 seq.) e RTE (3 seq.)
Rex 5 (1 seq.)	L2 (14 seq.)
Rex 6 (1 seq.)	L1(7 seq.) e TX1(3 seq.)
Rex6-1_OL (1 seq.)	L1(2 seq.) e TX1 (6 seq.)

O grupo L1 e o grupo Tx1 formam um *cluster* bem suportado (99%) e ambos estão presentes no genoma dos anfíbios analisados (Figura 1 e 2), e inicialmente foram recuperados através de duas *queries* de Rex6 no genoma de *P. carvalhoi*. Quatro *clusters* que estão presentes no cluster L1 (92%), com confiança de 53%, 96%, 82% e 52% respectivamente, compostos apenas pelos TEs de anuros. Os elementos de *Oryzias latipes* (swimmer AF055640) e *Homo sapiens* (L1Human U93574- que dá nome ao agrupamento), pipa216133, L1AShc1 e L1AShc7, não se enquadram em nenhum dos quatro grupos, apresentando-se como uma politomia no *cluster* L1. No *cluster* Tx1, suportado por um *bootstrap* de 96%, pode-se notar apenas sequências de anuros, que agrupam com o elemento Tx1 de *X. tropicalis* (Tx1XT M26915- que dá nome ao grupo). Porém, sabe-se que elementos desse clado também estão presentes em cnidários, ascídias e peixes (PUTNAM *et al.*, 2007) Nesses dois *clusters* (L1 e Tx1), podemos ver que as sequências em geral apresentam maior divergência em relação a outros clados recuperados, como pode ser evidenciado pelos ramos mais longos na árvore filogenética (Material complementar no Anexo IV).

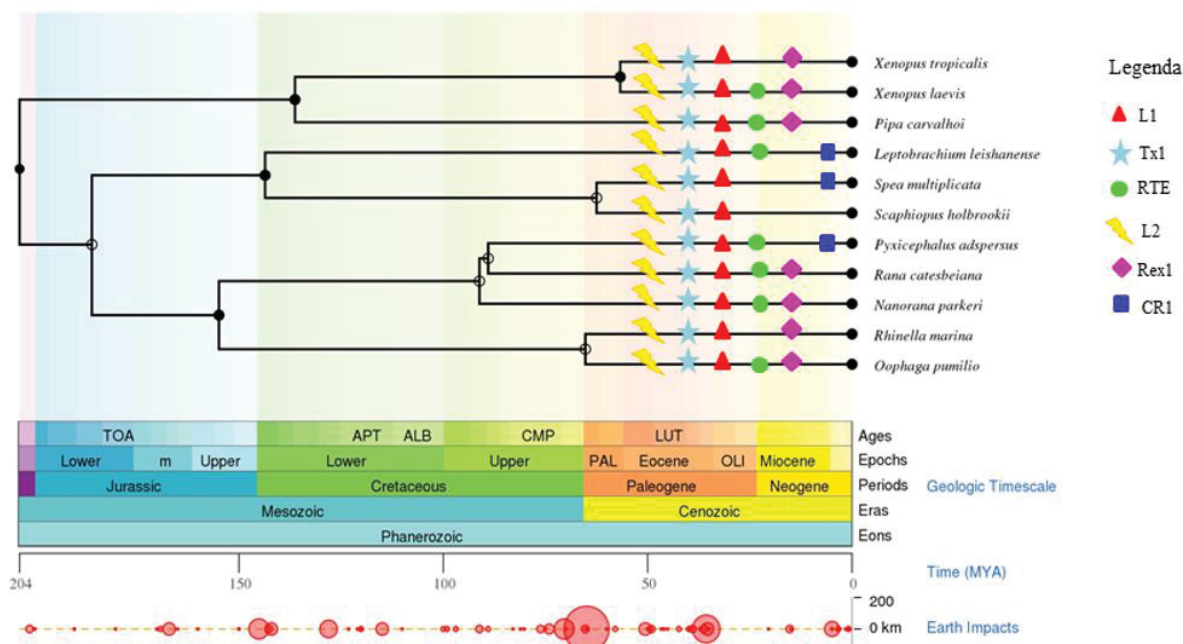


Figura 2: Cladograma contendo as espécies de anuros com genomas sequenciados e incluídas em nossas análises ilustrando os dados de presença/ausência das principais subfamílias/clados de TEs da ordem LINE (ver legenda). As relações filogenéticas entre as espécies foram compiladas a partir da plataforma *Timetree* (<http://www.timetree.org/book>) para a construção do cladograma. Círculos preenchidos representam nós inferidos com alta capacidade de refletir o real relacionamento filogenético entre os grupos, enquanto círculos não-preenchidos representam nós cujo relacionamento entre as espécies fora criado artificialmente durante a resolução das politomias observadas, de acordo com Hedges *et al* (2015).

No *cluster* RTE (54% de confiança), as sequências são provenientes dos genomas dos anuros *P. adspersus*, *L. catesbeiana*, *O. pumilio*, *L. leishanense*, *X. laevis*, *N. parkeri* e *P. carvalhoi* e agrupadas com elementos do genoma de mosquito *Aedes aegypti* (JAM1 Z86117), *C. elegans* (RTE1), sanguessuga *Helobdella robusta* (RTE5HRo) e com Rex3 de *Xiphophorus*. Elementos desse clado também estão presentes em plantas cnidários, nematoides, planárias e mamíferos (MALIK; EICKBUSH, 1998).

A sequência da sanguessuga (RTE5 HRo) ficou posicionada dentro do *cluster* de sequências dos anuros, agrupado especificamente com as cópias *P. carvalhoi*, com alto suporte (99%). Essa situação nos dá indícios para sugerir uma possível TH, como fonte de tal similaridade entre essas sequências de espécies filogeneticamente pouco relacionadas. Em um cenário de transferência vertical (TV) esperaríamos que as cópias provenientes de *P. carvalhoi* agrupassem com as cópias provenientes de outros anfíbios, principalmente com as espécies pertencentes à família Pipidae (*X. tropicalis* e *X. laevis*). Da mesma forma, seria esperado o agrupamento da sequência de *H. robusta* com a sequência de *Eisenia fetida*, uma

espécie relacionada da mesma classe. A busca no genoma de *X. tropicalis* não recuperou cópias para esse elemento e a cópia de *X. laevis* não agrupa com a sequência de *P. carvalhoi*. Essa observação pode indicar que o elemento entrou em processo de degeneração após a separação dos gêneros *Pipa* e *Xenopus* e foi completamente perdido em *X. tropicalis*. Perda secundária também pode explicar a ausência desse elemento em outras espécies de anuros investigados (Figura 2). A baixa amostragem de sequência de elementos pertencentes a distintos grupos taxonômicos também poderia explicar esse agrupamento de sequências de táxons tão distantes, porém, mesmo assim, o agrupamento da sequência de *H. robusta* dentro do clado de *P. carvalhoi* não seria esperado. Inconsistências nas relações filogenéticas de elementos pertencentes a esse clado já foram relatadas anteriormente na filogenia de peixes, por Volff *et al*, (2001), onde os autores sugerem que pode ter ocorrido recorrentes eventos de TH ao longo da história evolutiva dos peixes, uma hipótese que o próprio autor julga como não muito parcimoniosa.

Inúmeros casos de transferência horizontal em eucariotos já foram relatados, (LUDWIG; VALENTE; LORETO, 2008; SILVA; MARGARET, 2000; WALLAU; VIEIRA; LORETO, 2018). Sabe-se que essa transferência pode ocorrer de parasitas generalistas ou de associações simbióticas, sendo que para que um TE seja trocado entre duas espécies é necessário o contato físico entre a espécie doadora e a espécie receptora, sendo necessário que as duas espécies ocupem o mesmo nicho ecológico (SILVA; LORETO; CLARK, 2004). Casos de parasitas da subclasse Hirudínea, à qual pertence *H. robusta*, já foram relatados na pele das espécies de *X. laevis* provenientes do Sul da África (TINSLEY, 1981), reforçando que as condições necessárias para TH podem ter existido.

Para avaliar essa possibilidade de TH, realizamos uma comparação da taxa de substituição sinônimas (dS) encontrada para as cópias do elemento das espécies potencialmente envolvidas na TH e a taxa encontrada para os genes constitutivos dessas espécies. A comparação de valores dS foi escolhida para a inferência de TH com base em outros trabalhos (BARTOLOMÉ; BELLO; MASIDE, 2009; LUDWIG; VALENTE; LORETO, 2008; SILVA; MARGARET, 2000). É conhecido que a taxa de dS é uma medida de evolução neutra na ausência de forte uso tendencioso de códons. Assim, no caso de TV do TE, é esperada a mesma proporção de substituições sinônimas ocorrendo nos TEs e nos genes constitutivos ao comparar duas espécies e os resultados devem ser avaliados também considerando o nível de uso tendencioso de códons tanto dos TEs quanto dos genes (LUDWIG; VALENTE; LORETO, 2008). O uso tendencioso de códons influencia na

magnitude de dS por uma correlação negativa (Shields *et al.* 1988; Sharp e Li 1989), ou seja, se o gene constitutivo apresenta um forte uso tendencioso de códons, a taxa de dS será menor do que o esperado para evolução neutra.

Os genes escolhidos para o teste apresentam um baixo uso preferencial de códons (usando entre 52 e 56 dos 61 códons) e similar ao encontrado para o RTE (usando 53 códons) (Tabela 5), sendo assim úteis para as comparações de dS propostas, e conseqüentemente, no caso de TH, a taxa dS do RTE deve ser menor que a dos genes constitutivos. Os valores de dS para os genes constitutivos variaram de 0,78 a 0,83, similares ao valor encontrado para o elemento (0,795) (Figura 3), sugerindo que o elemento foi herdado de maneira vertical do ancestral comum, juntamente com os genes constitutivos. Alternativamente, um evento muito antigo de TH pode ter acontecido entre espécies ancestrais de modo que não temos mais capacidade de identificar, pois a ocorrência de substituições múltiplas, que podem ter acontecido nos genes constitutivos desde a separação dos ancestrais de *Xenopus* e *Helobdella*, podendo ter subestimado o valor real de dS. Outros eventos ao longo da evolução poderiam explicar as inconsistências observadas, como perda estocástica, polimorfismo ancestral ou diferentes taxas evolutivas (CAPY; ANXOLABÉHÈRE; LANGIN, 1994).

Tabela 5 Valores estimados de número efetivo de códons (ENC) e índice de códon *bias* (CBI), calculados no DAMBE para cada gene constitutivo e do RTE comparando *X. tropicalis* e *H. robusta*.

Identificação	Nome	ENC (21-61)	CBI (0-1)
Gene 1	Armadillo repeat containing 3	55,28	0,23
Gene 2	Glutamine-dependent NAD(+) synthetase	56,57	0,22
Gene 3	Glutamate-cysteine ligase	55,28	0,22
Gene 4	Hypoxia up-regulated protein 1	54,75	0,22
Gene 5	Dual serine/threonine and tyrosine protein kinase	55,46	0,23
Gene 6	Rab-GTPase-TBC domain	53,45	0,22
Gene 7	rspry1	54,06	0,29
Gene 8	Protein SDA1 homolog	52,85	0,23
RTE	RTE	53,39	0,30

Nota: ENC varia de 21, sendo máximo uso tendencioso de códons, até 61, quando não existe uso tendencioso de códons. CBI varia de 0, na ausência de uso tendencioso de códons, a 1, na presença de uso tendencioso máximo.

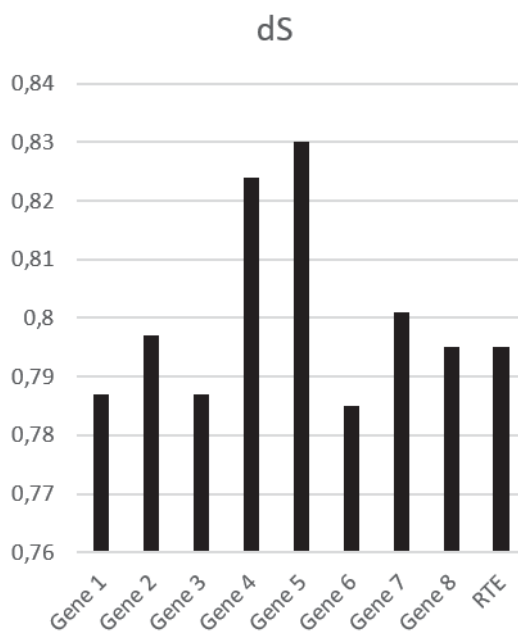


Figura 3: Estimativa da taxa de substituições sinônimas (dS) para genes constitutivos, conforme listado na tabela 4 e do elemento RTE (pipa84207 e RTE5_HRo).

As cópias de Rex1 recuperadas no genoma de *P. carvalhoi* agruparam em um *cluster* bem suportado (96%) com elementos provenientes de peixes, anfíbios e répteis. Dentro desse grupo, as sequências de *P. carvalhoi* se subdividiram em três agrupamentos: (1) agrupando duas sequências de *P. carvalhoi* com Rex1-4 de *X. tropicalis* e Rex1-6 de *Danio rerio*; (2) o segundo agrupa várias sequências de *P. carvalhoi* com os Rex1 provenientes de *X. laevis* (Rex1Xlc6L), *D. rerio* (Rex1-31, Rex1-4 e Rex1-8), *R. marina* (Rex1Rmc1), *L. catsebeianus* (Rex1Lcac1) e *X. tropicalis* (Rex 1 XT) e em um último *cluster*, agrupam cópias de *P. carvalhoi* com elementos provenientes de peixes e *X. laevis*. Esses clados são relacionados e incluem ainda sequências de outros peixes como *Oryzias latipes*, *Takifugu rubripes*, *Gasterosteus aculeatus*, *Salmon salar*, *Petromyzon marinus*, e do lagarto *Anolis carolinensis*. Com exceção de uma cópia denominada *pipa 771311* que apresentou uma maior divergência em relação às demais sequências (como pode ser observada pelo grande comprimento de ramo -Anexo IV), várias sequências reunidas nesse agrupamento apresentam baixa divergência entre si, indicando atividade mais recente de transposição. A existência de um *cluster* contendo 14 sequências de *P. carvalhoi* pode indicar um evento de *burst* de retrotransposição nessa espécie.

Esses agrupamentos sugerem que invasão do elemento Rex1 nesses genomas tenha sido um evento antigo, com cópias já presente no ancestral dos Gnathostomata (vertebrados

com mandíbula) e Agnatha (vertebrados sem mandíbula), uma vez nesse *cluster* agrupam também elementos provenientes do genoma de um peixe Agnatha popularmente denominado lampreia (*Petromyzon marinus*). Além disso, a distribuição de sequências das mesmas espécies em diferentes clados indica a presença de diversidade (possivelmente diferentes famílias) já no ancestral dos vertebrados e diversas ondas de retrotransposição de Rex1 devem ter ocorrido em diferentes linhagens, como em *P. carvalhoi*. Isso já foi demonstrado também para espécies de peixes (VOLFF; KORTING; SCHARTL, 2000). Contudo, cópias dos TEs pertencentes a esse *cluster* estão presente em apenas alguns genomas de anuros, indicando que esses elementos podem ter sido eliminados independentemente do genoma de algumas linhagens, o que já foi reportado também para algumas espécies de peixes, como por exemplo Celacanto (*Latimeria chalumnae*) (CHALOPIN *et al.*, 2013). Uma vez que detectamos grande diversidade de elementos Rex1 nos genomas de anuros e observamos sua presença já no ancestral dos peixes, perdas desses elementos é a hipótese mais parcimoniosa para explicar tal cenário.

O segundo *cluster*/superfamília com maior agrupamento de sequências recuperadas de *P. carvalhoi* foi o L2 (*bootstrap* 68%), onde também estão presentes cópias dos elementos de todas as espécies de anuros analisadas, além de elementos provenientes de peixes (MauiAF086712 de *Fugu rubripes* e Rex-5 *Xiphophorus*) e do urocordado *Ciona intestinalis* (L2-like AABS01001177). A distribuição de L2 foi inicialmente descrita em cnidários, vertebrados e insetos (LOVŠIN; GUBENŠEK; KORDIŠ, 2001). As sequências de anuros foram recuperadas em dois agrupamentos distintos, o primeiro composto das cópias de representantes da família Pipidae (*P. carvalhoi*, *X. tropicalis* e *X. laevis*) (98% de suporte) e o segundo com os demais anuros analisados. Os elementos de peixe agruparam entre si. Assim como no clado Rex 1 as cópias de *P. carvalhoi* presentes em L2 apresentam uma divergência menor comparada com os outros *cluster*, como indicado pelo menor tamanho dos ramos (Figura Anexo IV), indicando que esse elemento, apesar de antigo em anuros, pode ter tido ondas recentes de retrotransposição.

Por fim, poucas cópias de elementos CR1 foram recuperadas dos genomas analisados, estando presente em apenas três espécie das 10 espécies analisadas de anuros incluídas em nosso estudo (*L. leishanense*, *P. adspersus* e *S. multiplicata*). Esse clado ainda inclui cópias de elementos recuperados do genoma do platelminto *Schistosoma mansoni* (SR1 U66331) e de *Gallus gallus* (CR1 U88211), do mosquito *Anopheles gambiae* (T1 M93689) e do nematódeo *Caenorhabditis elegans* (Sam6 Z82275). Apesar de nossas estratégias de

amostragem não serem designadas para avaliar o clado CR1, análises anteriores (CHALOPIN *et al.*, 2013), demonstraram que esse elemento é amplamente distribuído e tem grande importância na evolução dos tetrápodes, juntamente com REX1 e L1, e estima-se que esses elementos foram preservados e proliferaram da transição aquática para terrestre em tetrápodes (CHALOPIN *et al.*, 2015; SHAO; HAN; PENG, 2019; SOTERO-CAIO *et al.*, 2017).

Nossos dados fornecem mais evidências de que as sequências genericamente denominadas Rex não correspondem a um grupo evolutivamente relacionado, uma vez que as cópias desses elementos não foram recuperadas como um único *cluster*, e sim distribuídas em diferentes *clusters* (Rex1 no *cluster* Rex1, Rex3 no *cluster* RTE, Rex5 no *cluster* L2 e Rex 6 no *cluster* R4). Essa estratégia de nomenclatura usando iniciais do tipo de retrotransposon, espécie e número de acordo com a ordem de descrição/descoberta é bastante utilizada, porém, pode gerar confusões, como no caso dos elementos Rex.

Ainda, a ausência de um sistema uniformizado de classificação das cópias de elementos de transposição tem historicamente gerado inconsistências na forma de descrição de novas famílias de elementos. Os dois principais artigos sobre classificação de TEs (KAPITONOV; JURKA, 2008; WICKER *et al.*, 2007) possuem grandes diferenças em relação aos elementos do tipo LINE e podem impactar a interpretação de resultados. Por exemplo, a classificação de Wicker *et al.*, (2007) tem máximo cinco superfamílias de LINE e não há informação dos clados que estão presentes em cada uma dessas superfamílias. A classificação de Kapitonov e Jurka (2009) apresenta um número maior de superfamílias, mas ainda não engloba todos os clados e também, não indica uma relação entre os clados.

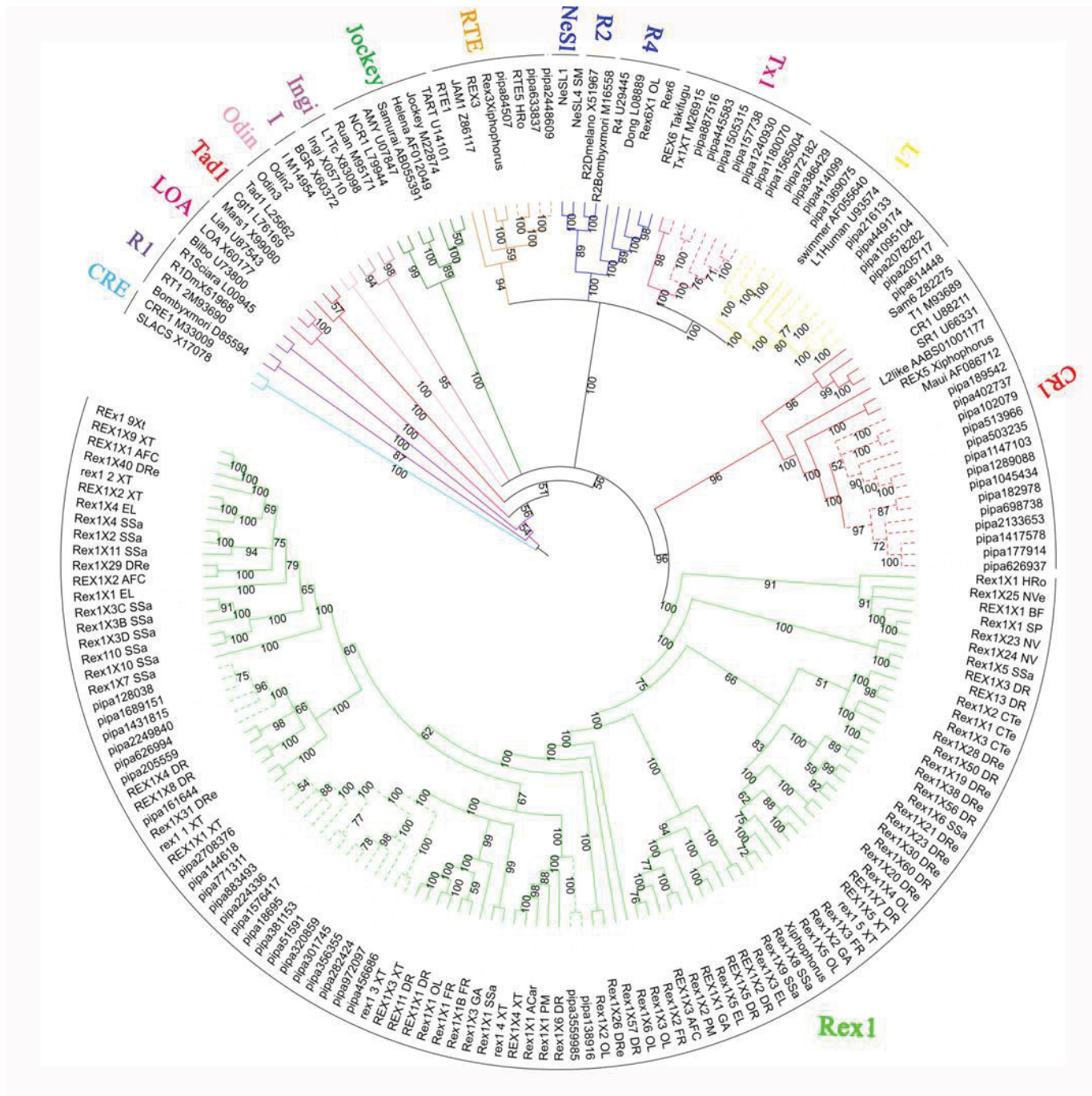
No nosso entendimento, Kapitonov *et al.* (2009) apresentam uma classificação mais completa para esses elementos mostrando não somente 28 *clusters* de LINEs, mas os agrupando em seis grandes grupos. Nesse sentido, as relações entre os *clusters* também são importantes, já que a distribuição dos *clusters* em grandes grupos deveria ser realizada com base nas relações evolutivas entre os elementos. Em nosso trabalho, um único grupo foi recuperado para os *clusters* RTE, NeSL, R2, R4, Tx1 e L1, o que não foi reportado anteriormente. Nossos dados refletem que a divisão proposta por Kapitonov e Jurka (2009) pode não refletir a origem e evolução dos elementos LINEs. Por exemplo, *clusters* considerados do grupo Jockey, Rex1, CR1 e L2 não agrupam com o clado Jockey. Ainda, R2, R4 e NeSL não agrupam com os elementos CRE, todos considerados do grupo R2. De fato, esse agrupamento de R2 e R4 com CRE não foi evidenciado em nenhum artigo. R1, Loa,

Tad1 e I (grupo I) também não formam um único grupo em nossos dados. Na literatura, a único *cluster* desse grupo também não é recuperada, já que o clado Jockey também agrupa com esses *clusters* (BEREZIKOV; BUCHETON; BUSSEAU, 2000; KOJIMA; FUJIWARA, 2005; NOVIKOVA *et al.*, 2007; NOVIKOVA; FET; BLINOV, 2009). Esses resultados mostram a complexidade da classificação dos elementos do tipo LINE e a necessidade de atualizar ou estabelecer uma nova classificação para esses elementos.

De forma geral, os dados apresentados nesse trabalho indicam que os elementos LINES recuperados para os anuros são antigos, provavelmente já estavam presentes no ancestral de peixes e anuros e representam um importante elemento na evolução dos genomas desse grupo, sendo que ondas de retrotransposição em linhagens específicas podem ter ocorrido independentemente durante a diversificação das espécies.

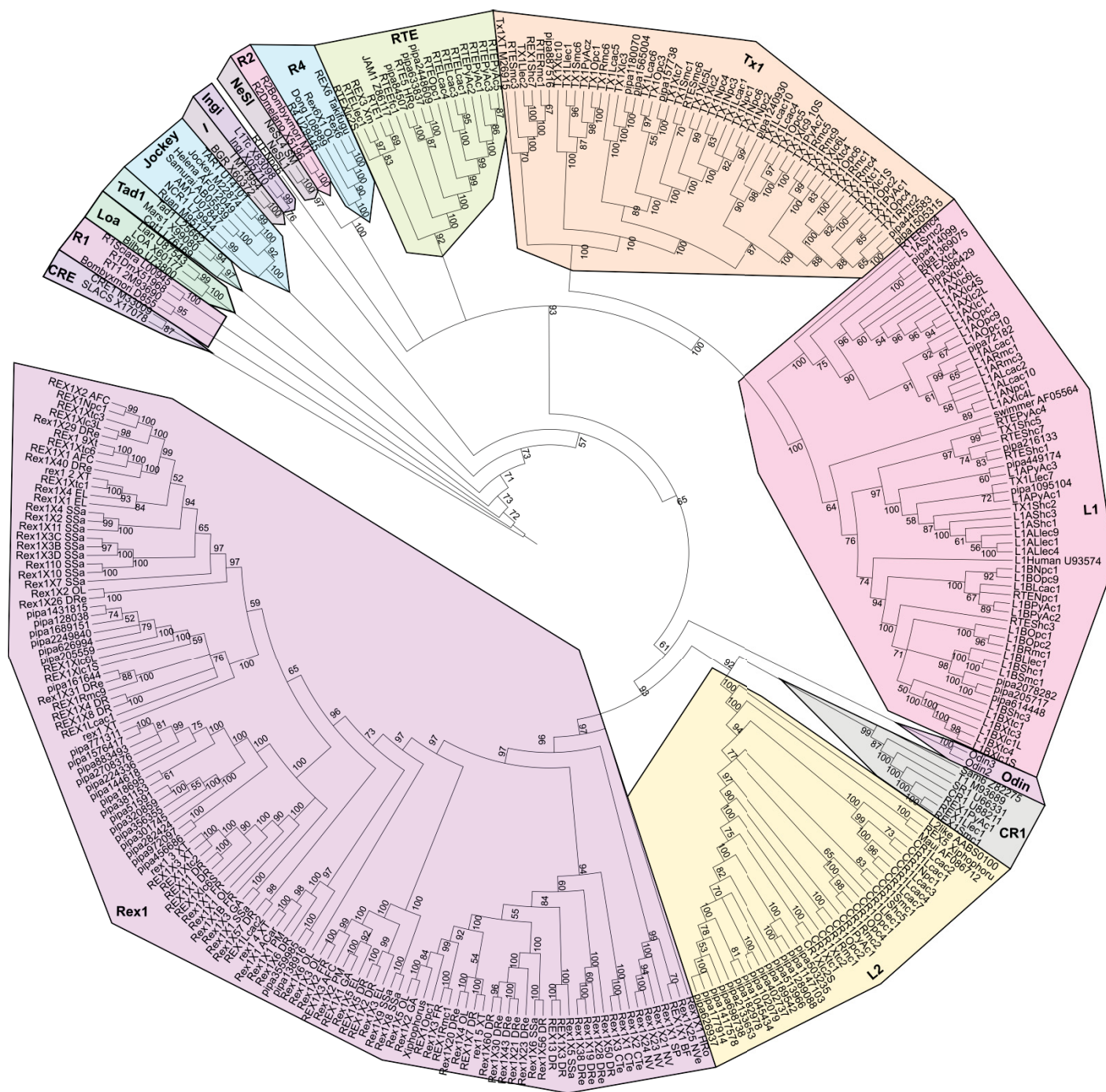
Anexo I- Árvore Bayesiana exploratória inferida com as sequências de *P. carvalhoi*

Árvore filogenética reconstruída a partir do domínio da RT dos elementos LINES recuperados no genoma de *P. carvalhoi*. Os valores de Probabilidade Posterior são indicados próximo aos nós. As linhas pontilhadas representam os contigs recuperados em *P. carvalhoi*



Anexo II- Árvore Bayesiana

Análise bayesiana (BA) implementada no MrBayes 3.2.6 (RONQUIST *et al.*, 2012), sob um modelo misto, 10,000,000 de gerações de cadeia de Markov Monte Carlo (MCMC), amostrando árvores a cada 1,000 gerações. Os 25% dos resultados iniciais, considerados “burn-in” foram descartados. Apenas os clados estão coloridos. As cópias de elementos dos anuros (exceto *P. carvalhoi*) estão nomeadas de acordo com a *query* usada para encontrá-las, porém, pode-se evidenciar que algumas pertencem a clados distintos do clado da *query* usada.



Anexo III- Descrição dos *contigs* analisados de *P. carvalhoi*, evidenciando em colorido os *contigs* utilizados na árvore filogenética.

Summary of the analyzes of the 15 most conserved copies of REX superfamily. The information indicated are: Contigs / scaffolds of each copy with initial and final position of the isolated region, size of each protein. Legend: *Res 1*, *L2*, *RT*, *L1* and *TK1*

Genome/Chromosome	ORF2				ORF1						
	bp/aa	start	stop	Domain	position	e-value	bp/aa	start	stop	Domain	ositiowalue
<i>Pipa carvalhoi</i>											
pipa402737	2919 972	451	3369	RT_nLTR_like	618-876	1.21e-36					
pipa513966	837 278	2185	3021	Exo_endo_phos	117-350	8.59e-04				degenerada	
pipa1417578 (missing)	669 222	1042	1710	RT_like super family (truncada)	225-278	8.50e-06				degenerada	
pipa102079	687 228	2354	3040	RT_nLTR_like	39-223	1.73e-34				Ausente	
pipa352038				RT_like super family (truncada)	39-142	1.62e-28	131 37	1196	2326	base_M78 super	1291-14.26e
pipa503235				degenerada (sem sequência de RT, logo não foi para filogenia)						degenerada	
pipa2133653	732 243	287	>1018	RT_nLTR_like (faltando na ponta) (traduzida no genedoc)	112-371	1.38e-40				degenerada	
pipa698738 (missing)	1131 376	253	>1383	RT_like super family (degenerada, faltando um pedaço no meio e na ponta)	6-181	7.34e-18				degenerada	
pipa1045434	597 198	1893	2489	RT_nLTR_like	39-297	4.59e-41				degenerada	
pipa626937	840 279	287	>1126	RT_like super family	6-181	5.62e-19				degenerada	
pipa189542	645 214	1835	2479	RT_like super family	6-182	3.89e-22				degenerada	
pipa1147103	855 284	221	1075	RT_like super family (degenerada, uma RT pequena)	38-191	5.54e-24				degenerada	
pipa1289088 (missing)	1299 432	80	>1378	super family (degenerado, faltando um pedaço da orf, porém o contig é p	8-154	1.09e-12				degenerada	
pipa177914	690 229	1679	2368	RT_nLTR_like (bem grande a orf, porém o contig acaba)	9-251	1.88e-31				degenerada	
pipa182978	1236 411	178	1413	RT_like super family (quebrada na ponta)	39-218	9.78e-29				degenerada	
pipa128038	5 144/609/233/161/467/>22			RT_nLTR_like	103-361	4.46e-44				degenerada	
pipa224336	1428 475	184	1611	RT_like super family (truncada) + DUF1891	(62-144 + 7-88)/ 94-135	(8.49e-18 + 1.15e-10) / 3.28e-10				degenerada	
pipa1576417	2 73/351/1188/638/409/988			RT_like super family + DUF1891	82-336/342-383	3.77e-45/				degenerada	
pipa51591	966 321	2734	3699	RT_like super family (truncada)	1-72/19-75	9.97e-14/ 8.57e-06				degenerada	
pipa1904528	558 185	961	1518	nLTR_like (faltando uma parte no meio, mas perto do começo)+ DUF18	1-171/177-217	5.25e-19/ 9.36e-06				degenerada	
pipa282424	4 77/633/2174/5044/07/1136			DUF1891	41-82	7.10e-11				degenerada	
pipa2708376 (missing)	915 304	84	>998	RT_like super family (truncada) + DUF1891	(1-77+ 1-61)/ 67-108	(7.76e-05 + 1.50e-05) / 5.88e-07				degenerada	
pipa232970	723 240	1943	2665	RT_like super family + DUF1891	1-225/231-272	7.20e-27/ 2.35e-05				degenerada	
pipa1118569	717 238	517	1233	RT_nLTR_like (faltando um pedaço no começo)+ DUF1891	7-88/95-135	2.32e-09/ 1.06e-07				degenerada	
pipa577703	897 298	1433	2329	RT_like super family (faltando pedaços nas pontas) + DUF1891	2-197/203-231	9.82e-20/ 1.43e-03				degenerada	
pipa381153	1329 442	963	2291	RT_nLTR_like (faltando pedaços nas pontas)+ DUF1891	1184-1945/ 1961-2080	3.88e-34/ 1.17e-03				degenerada	
pipa626994	1224 407	2655	3878	Exo_endo_phos	122-323	1.32e-04					
pipa913683	831 276	997	1827	RT_nLTR_like + DUF1891	79-332/ 338-379	3.97e-31/ 7.76e-08				degenerada	
pipa972097	1062 353	2644	3705	RT_nLTR_like (degenerada) + DUF1891	130-383/389-407	4.84e-41/ 4.54e-04				degenerada	
pipa382562	810 269	4951	5760	RT_like super family + DUF1891	14-125/ 131-172	2.77e-14/ 1.09e-04				degenerada	
pipa480622 (degenerada)	519 172	3092	3610	RT_nLTR_like (degenerada)	111-340	5.88e-36				degenerada	
pipa144618	672 223	3852	4523	RT_nLTR_like (faltando na ponta e no contig) + DUF1891	3-172/178-219	7.39e-15/ 3.89e-06				degenerada	
pipa169719	630 209	4782	5411	RT_nLTR_like (faltando um pedacinho na pontinha)	30-67	1.11e-03					
pipa456686	1800 599	6332	>8131	DUF1891	115-212	2.46e-29				degenerada	
pipa200596	1800 599	6332	>8131	RT_nLTR_like	68-109	1.53e-04					
pipa300574	987 328	2321	3307	RT_nLTR_like	376-600	9.86e-37					
pipa138916	1338 445	4099	>5436	RT_like super family (faltando na ponta e no meio) + DUF1891 super fami	1-172/178-219	6.10e-14/ 3.23e-04					
pipa3559985	489 162	9630	10118	RT_nLTR_like	173-427	1.64e-41					
pipa18695	1029 342	8254	9282	RT_like super family (faltando nas pontas)	37-156	6.32e-22					
pipa123463	762 253	16	>777	RT_nLTR_like (degenerado)	115-342	9.66e-41					
pipa123463	762 253	16	>777	RT_nLTR_like	29-254	4.14e-46					
pipa123463	855 284	18	>872	RT_nLTR_like	60-285	9.81e-45					
pipa123463	1329 442	963	2291	Exo_endo_phos	122-323	1.32e-04					
pipa123463	1449 482	2502	3950	RT_nLTR_like + DUF1891	79-332/ 338-379	3.97e-31/ 7.76e-08					
pipa123463	507 168	3260	3766	RT_like (faltando no meio e nas pontas)	2-154	4.73e-21				degenerada	
pipa123463	561 186	267	827	Exo_endo_phos	19-144	3.93e-03					
pipa123463	372 123	2193	2564	RT_like super family (degenerada na ponta) + DUF1891 super family	6-62/ 68-109	2.03e-05/ 5.40e-04					
pipa123463	1014 337	609	1622	RT_like (Reverse transcriptase (truncada))	120-213 + 6-92	8.17e-14 + 3.52e-08				degenerada	
pipa123463	173 + 351	731/11252/46		Exo_endo_phos	54-231	1.98e-04				degenerada	
pipa123463	173 + 351	731/11252/46		RT_like (Reverse transcriptase (truncada))	116-173/ 3-96	4.65e-12/ 1.35e-04					

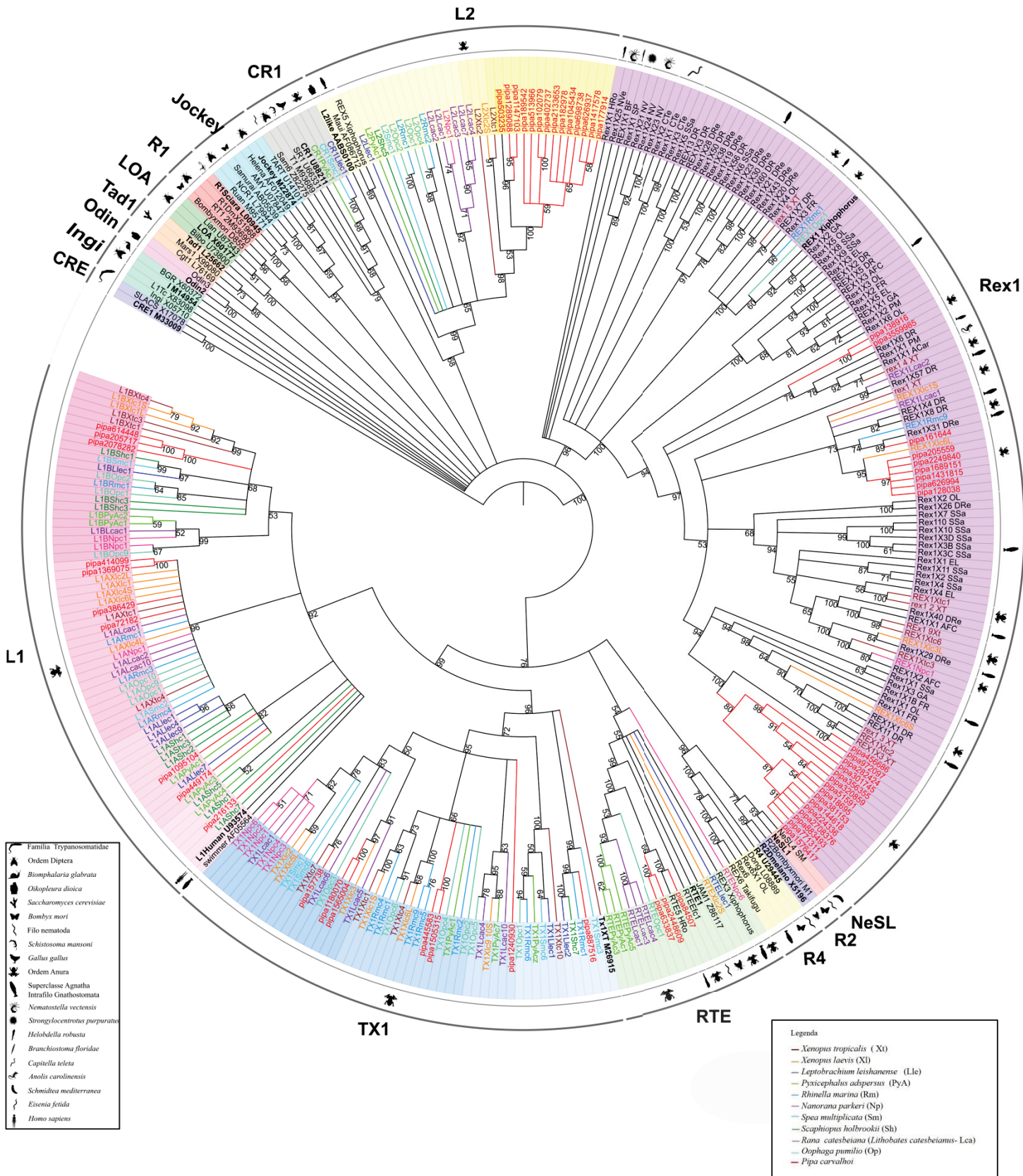
pipa305943	264 87	3988	4251	Exonuclease-Endonuclease-Phosphatase (EEP)	(6-)87	6.35e-18									
	465 154	6245	6709	RT like (degenerado NAS PONTAS E NO MEIO)	3-152	4.58e-17									degenerada
pipa1710062	1617 538	1462	3078	Exonuclease-Endonuclease-Phosphatase (EEP) + RT like (degenerado)	23-154/ 452-535	9.18e-03/ 7.18e-18									degenerada
pipa1431815	144 + 204	785/102219/	122	RT like: Reverse transcriptase (truncada)	62-144 + 7-61	9.21e-19 + 6.04e-08									
pipa416182	693 230	3974	>4666	like super family (degenerada nas pontas e no meio) + DUF1891 super fa	1-172 / 178-219	2.72e-18/ 2.23e-04									
pipa1021816 (só ap	642 213	1628	2269	RT like super family (degeneradae truncada)	1-172	2.15e-17									
pipa1134641	909 302	2463	3371	RT like super family (degeneradae truncada)	176-280	7.43e-24									
pipa51591	966 321	2734	3699	like super family (degenerada na ponta e no meio) + DUF1891 super fat	1-171/ 177-217	5.25e-19/ 9.36e-06									
pipa899637	450 149	1253	1702	Exonuclease-Endonuclease-Phosphatase (EEP) (degenerado)	16-110	2.42e-03									
	633 210	2765	3397	RT like (na ponta degenerado)	54-172	6.39e-13									
pipa205559	600 199	910	>1509	RT like	2-198	9.04e-26									
pipa356355	1116 371	1749	2864	RT_nLTR like	115-368	115-368 /									
pipa300262	387 128	1054	1440	Exonuclease-Endonuclease-Phosphatase (EEP) (acho que completa)	16-112	7.47e-03									
pipa96671	876 291	342	1217	Exonuclease-Endonuclease-Phosphatase (EEP) (degenerado)	(7)-95	8.15e-03									
pipa43444	909 302	2231	3139	Exonuclease-Endonuclease-Phosphatase (EEP) (degenerado)	19-148	8.81e-05									
pipa1071858				SEM DOMÍNIO											
pipa1150269	858 285	53	910	Exonuclease-Endonuclease-Phosphatase (EEP)	15-147/	1.90e-06/									
pipa144618	1800 599	6332	>8131	RT_nLTR like	376-600	9.86e-37									
pipa320859	951 316	2766	>3716	RT_nLTR: Non-LTR	47-298	8.68e-43									
pipa237640				SEM DOMÍNIO											
pipa2297285	852 283	7	858	Exonuclease-Endonuclease-Phosphatase (EEP) (degenerado)	3-133/ 90-186	3.73e-04/ 1.50e-03									
	225 74	1155	>1379	ke: Reverse transcriptase (quebra nas pontas, talvez por estar no final do c	54-75	8.91e-05									
pipa1037303				SEM DOMÍNIO											
pipa301745	792 263	3455	>4246	RT like: Reverse transcriptase	54-250	3.35e-25									
pipa48022	399 132	7880	8278	RT like: Reverse transcriptase(degenerada, faltando pedaços nas duas pont	1-115	1.49e-06									
pipa883493	1032 343	887	1918	RT_nLTR: Non-LTR	82-334	3.76e-41									
pipa130056	258 85	3073	3330	RT like: Reverse transcriptase (quebrada na ponta)	(7)-61	2.36e-07									
pipa96843	1008 335	3838	>4845	RT like super family (quebrada na ponta)	261-336	7.73e-18									
pipa31988	588 195	1817	2404	DUF1891	50_91	2.02e-04									
	402 133	1432	1833	DUF1891	1-132	1.71e-12									
pipa1689151	672 223	778	1449	as não achei o outro domino de RT, apenas o DUF, mas no cd serach mos	51-212	1.94e-25									
pipa287655	345 114	1193	1537	RT like: Reverse transcriptase (uma das pontas quebradas)	(07)-1988	1.53e-11									
pipa146549	987 328	719	1705	RT like: Reverse transcriptase (ponta quebrada)	169-273	1.03e-21									
pipa90867	657 218	2129	2785	RT like: Reverse transcriptase (ponta quebrada)	54-218	5.08e-20									
pipa771311	163 + 456 8 + 1049 + >14			Rt truncada	40-141 + 6-62	1.42e-29 + 1.25e-0									
pipa38975	642 213	3466	4107	RT like super family (degenerada na ponta) + DUF1891 super family	1-172 / 178-206	5.87e-17/ 5.73e-03									
pipa1044384	633 210	1526	2158	like super family (degenerada na ponta e no meio) + DUF1891 super fan	11-88/ 94-135	1.93e-09/ 1.16e-10 5.96e-20/ 6.83e-10									
pipa162327	744 247	3819	4562	like super family (degenerada na ponta e no meio) + DUF1891 super fan	22-192/ 198-239										
pipa76109	327 108	673	999	RT like super family (degenerada na ponta) +	19-67	8.93e-04									
pipa650303															
pipa69139	155 + 354 25 + 4192 + 451			RT like: Reverse transcriptase (truncada, só achei duas partes)	57-152+ 6-62	2.44e-22+ 3.03e-04									
pipa650303	8 65+ 249 99 + 18496 + 21				27-61 + 2-47	3.72e-04 + 1.80e-03									
pipa288515				sem DOMÍNIO											
pipa130599	579 192	4074	4652	DUF1891	43-83	1.83e-03									
pipa3243310	819 272	<3	>821	like super family (degenerada na ponta um pedacinho) + DUF1891 super f	1-207/213-254	3.43e-29/ 9.74e-11									
pipa3120682				Sem DOMÍNIO											
pipa843200	333 110	1329	1661	RT like super family	36-92	7.01e-04									
pipa2455845				Sem DOMÍNIO											
pipa2165971				Sem DOMÍNIO											

pipa937612					Sem DOMÍNIO														
pipa451576	372	123	175	>546	RT like super family (truncada) + DUF1891 super family	9-62/68-109		9.06e-05/7.43e-06											
pipa303719 ATG e c	47 + 678	19 + 3762	104		RT like super family (degenerada na ponta)	34 + 25-105 / 111-15		1.37e-04 + 1.19e-08 / 1.27e-05											
pipa32610	258	85	766	1023	RT like super family (degenerada na ponta)	(4)-61		4.87e-05											
pipa450055	975	324	6391	7365	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	1-172/ 178-219		9.67e-14/ 6.00e-03											
pipa2336844	510	169	69	>578	DUF1891 super family	43-84		3.38e-05											
pipa1683348	312	103	233	544	RT like super family (quebrada no meio e nas pontas)	(2)-45		6.15e-03											
pipa310420	654	217	2204	2857	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	6-62/ 68-109		5.84e-04/ 1.69e-03											
pipa728354	294	97	459	752	DUF1891 super family	54-95		1.22e-04											
pipa714496	318	105	846	1163	RT like: Reverse transcriptase	19-75		6.20e-04											
pipa436291	315	104	420	>734	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	3-311/ 327-452		1.69e-15/ 1.21e-09											
pipa867354	978	325	22	999	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	1-170/ 176-217		4.20e-17/ 7.61e-05											
pipa594658	369	122	849	1217	RT like: Reverse transcriptase (quebrada no meio e nas pontas)	119-228		1.39e-26											
pipa312347	852	283	435	1286	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	17-127/ 134-175		1.26e-17/ 1.51e-05											
pipa31704	801	266	367	1167	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	7-115/ 121-162		1.07e-16/ 3.41e-04											
pipa906848					SEM DOMÍNIO														
pipa60052	711	236	1782	2492	RT like: Reverse transcriptase	(9)-195		2.23e-17											
pipa632243	522	173	1438	1959	DUF1891 super family	54-95		4.82e-05											
pipa1955760	471	156	231	701	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	7-88/ 94-135		7.23e-12/ 1.41e-10											
pipa589684	735	244	222	956	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	3-97/103-144		3.22e-10/ 9.76e-10											
pipa14760	717	238	3044	3760	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	94-135/7-88		4.79e-10/ 2.00e-08											
pipa2425106					sem DOMÍNIO														
pipa279164	714	237	1259	1972	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	6-88/94-135		4.56e-11/ 7.11e-11											
pipa594690					SEM DOMÍNIO														
pipa1859884	495	164	139	633	DUF1891 super family	20-61		1.78e-08											
pipa241134	228	75	400	>627	DUF1891 super family	41-76		1.24e-10											
pipa1939599	339	112	172	>510	RT like: Reverse transcriptase	2-100		5.73e-13											
pipa1115245	426	141	217	>642	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	7-88/94-135		3.17e-10/ 94-135 1.16e-11											
pipa249184	180	59	5056	5235	DUF1891 super family (quebrada)	41-59													
pipa161644	786	261	8	793	RT like super family + DUF1891 super family	1-226/232-253		2.97e-39/ 1.44e-03											
pipa3050503					sem dOMÍNIO														
pipa2869788	387	128	267	>653	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	7-85/91-128		4.78e-14/ 1.11e-08											
pipa3069408	360	119	148	507	RT like super family (quebrada no meio e nas pontas)	15-95		3.09e-11											
pipa2449107	456	151	122	>577	like super family (quebrada no meio e nas pontas) + DUF1891 super fam	1-111/117-152		1.39e-22/ 1.79e-09											
pipa231414					sem dOMÍNIO														
pipa314492	160 + 369	81 + 106	63 + 136		RT like+RVT 1	128-152 + 1-99		1.37e-03 + 1.19e-04											
	579	192	1405	1983	DUF1891 + PKc like	43-84 + 102-172		7.26e-06 + 7.26e-03											
pipa2061954					sem dOMÍNIO														
pipa787145					sem dOMÍNIO														
pipa536175					sem dOMÍNIO														
pipa1073759	525	174	123	>647	DUF1891 super family	41-82		1.01e-10											
pipa25747	675	224	2835	3509	RT like super family (quebrada no meio e nas pontas)	62-167		4.71e-28											
pipa3185421					sem DOMÍNIO														
pipa191159					sem DOMÍNIO														
pipa2249840	6	91 + 195	56 + 116	1 + 135	RT like super family (truncada) + DUF1891 super family	2-75 + 7-64		0.39e-10 + 2.93e-05											
pipa847362					sem DOMÍNIO														

pipa150317	648 215	209	856	RT like super family (truncada) + DUF1891 super family	10-65/71-112	5.65e-03/1.79e-08							
pipa1601637	339 112	241	579	RT like super family (quebrada no meio e nas ponts)	(6)-88	3.82e-13							
pipa526823				sem DOMÍNIO									
pipa134384	942 313	1051	1992	RT like super family (quebrada no meio e nas ponts)	1-170	6.98e-18							
pipa305791	507 168	3202	3708	RT like super family (quebrada no meio e nas ponts)	1-163	5.97e-23							
pipa209325	720 239	262	981	RT like super family (truncada) + DUF1891 super family	6-59/65-106	4.85e-04/1.06e-04							
pipa730287	792 263	333	>1124	RT like super family (quebrada no meio e nas ponts)	189-264	9.28e-17							
pipa390043	543 180	4561	5103	RT like super family (quebrada no meio e nas ponts)	125-168	1.65e-06							
pipa1347626	330 109	974	1303	RT like super family (quebrada no meio e nas ponts)	14-97	3.42e-12							
pipa84507	1302 433	4645	>5946	RT nLTR: Non-LTR	109-358	4.11e-46							
	762 253	3731	4492	Endonuclease domain (truncada)	3-154	7.26e-36							
pipa30089	396 131	2724	3119	RT like: Reverse transcriptase (quebrada no meio e nas ponts)	75-131	3.38e-06							
pipa1532542	558 185	1078	1635	RT like: Reverse transcriptase (quebrada no meio e nas ponts)	1-184	1.54e-28							
pipa633837	1068 355	1751	2818	RT nLTR: Non-LTR	91-343	8.69e-53							
pipa171948	579 192	1252	1830	RT like super family (quebrada no meio e nas ponts)	136-192	1.75e-12							
pipa75656	375 124 +	374	748	ranscriptase / RT like: Reverse transcriptase + group II intron reverse tra	-120/ 5-109 + 29-100	5.04e-17/ 1.42e-09 + 1.59e-04							
pipa72182	1173 390	54	1226	Exonuclease-Endonuclease-Phosphatase (EEP)	9-237	4.40e-26							
	1344 447	1442	2785	RT nLTR: Non-LTR (long terminal repeat)	47-305	1.32e-58							
pipa55565	492 163	4439	4930	RT nLTR: Non-LTR (pontinha degenerada)	23-122	1.64e-31							
pipa216133	2562 853	847	3408	Endonuclease domain (L1-EN)/ RT nLTR: Non-LTR	14-230/507-764	2.94e-39/1.57e-56							
pipa587510	516 171	1202	1717	RT like: Reverse transcriptase (quebrada na pontinha)	14-97	7.00e-13							
pipa449174	1845 614	1439	3283	RT nLTR: Non-LTR	204-465	1.06e-56							
pipa1240930	1233 410	228	1460	Arginine methyltransferase-interacting protein, contains RING Zn-finger	140-242	1.62e-07							
	3369 1122	1457	4825	Endonuclease domain/RT nLTR: Non-LTR	5-232/ 495-751	7.17e-37/5.86e-56							
pipa1209579	330 109	152	481	ranscriptase / RT like: Reverse transcriptase + group II intron reverse tra	1-95/16-88	7.22e-15/5.61e-03							
pipa2777315	444 147	113	556	RT like: Reverse transcriptase (quebrada no meio e nas ponts)	93-147	8.29e-14							
pipa2448609	183 708	66 + 88	17 + 159	RT like: Reverse transcriptase (truncado)	113-183 + 19-126	8.84e-19 + 2.92e-13							
pipa445583	3345 1114	1967	5311	Endonuclease domain/ chromosome segregation protein/ RT nLTR: Non-LTR /zinc-binding in reverse transcriptase:	6-234/287-449/ 497-	5.71e-31/ 3.32e-05/ 6.41e-55/ 1.74e-03							
	858 285	478	1335	universal minicircle sequence binding protein	185-245	7.91e-12							
pipa1180070	1764 587	10	>1773	RT nLTR: Non-LTR	268-523	3.56e-60							
pipa414099	669 222	360	1028	RT nLTR like	1-204	5.53e-41							
pipa205717	1425 474	1084	2508	RT nLTR: Non-LTR	104-364	1.46e-52							
pipa614448	1116 371	82	>1197	RT nLTR: Non-LTR	104-364	4.90e-61							
pipa386429	735 244	118	852	Exonuclease-Endonuclease-Phosphatase (EEP)	18-114	3.81e-10							
	236 + 513	886 + 1	1596 + 2	RT nLTR: Non-LTR	130-235 + 14-128	5.77e-32 + 2.04e-18							
pipa1369075	177 + 270	90 + 19	23 + >22	RT nLTR: Non-LTR	10-173 + 1-67	1.09e-31 + 1.70e-07							
pipa1505315	2886 961	1643	>4528	Endonuclease domain (L1-EN) / RT nLTR: Non-LTR (6-234/497-755	4.57e-30/ 5.24e-55							
	1488 495	150	1637	universal minicircle sequence binding protein (UMSBP); Provisional	185-245	7.42e-11							
pipa52406	846 281	136	981	RT nLTR: Non-LTR (quebrada na pontinha)	104-275	2.59e-37							
pipa887516	819 272	1137	1955	RT nLTR: Non-LTR	4-260	1.25e-49							
pipa284208	843 280	1727	2569	RT nLTR: Non-LTR (quebrada na pontinha)	89-262	2.04e-40							
pipa2078282 (apare	570 189	643	>1212	RT nLTR: Non-LTR	1-190	1.46e-35							
pipa940990	1554 517	1130	2683	RT nLTR: Non-LTR (quebrada na pontinha)	297-504	3.97e-45							
pipa88930	966 321	232	1197	RT nLTR: Non-LTR (nas ponts)	153-321	7.86e-39							
pipa117387	2556 851	636	3191	Endonuclease domain / RT nLTR: Non-LTR	4-228/ 491-747	1.48e-32/ 4.14e-52							
	633 210	4	636	Arginine methyltransferase-interacting protein	(3)-61	5.37e-12							
pipa1095104	1008 335	142	1149	RT nLTR: Non-LTR/ group II intron reverse transcriptase/maturase	64-325/ 152-219	2.78e-54/ 4.85e-03							
pipa1565004	1116 371	168	>1283	RT nLTR: Non-LTR	64-319	1.12e-60							
pipa256168	1026 341	94	1119	RT nLTR: Non-LTR (quebrada na pontinha)	153-316	4.57e-35							
pipa1532542	558 185	1078	1635	e: Reverse transcriptase (quebrada no meio e nas ponts) provavelmente tr	1-184	1.54e-28							
pipa157738	2760 919	209	2968	Phosphatase (EEP) domain superfamily/ (RecF/RecN/SMC N terminal don	154/ 170-379/ 422-6	6.25e-19/1.75e-03/ 2.08e-47							
pipa1193139	492 163	445	936	RT nLTR: Non-LTR (pontinha quebrada)	47-152	3.79e-33							
pipa921022	777 258	636	1412	RT nLTR: Non-LTR (quebrada na ponta)	84-249	5.06e-40							
pipa387411	294 97	1499	1792	RT like: Reverse transcriptase (quebrada no meio e nas ponts)	(2)-81	2.24e-09							
pipa277531	771 256	284	>1054	RT nLTR: Non-LTR (quebrada na pontinha)	47-253	1.07e-45							

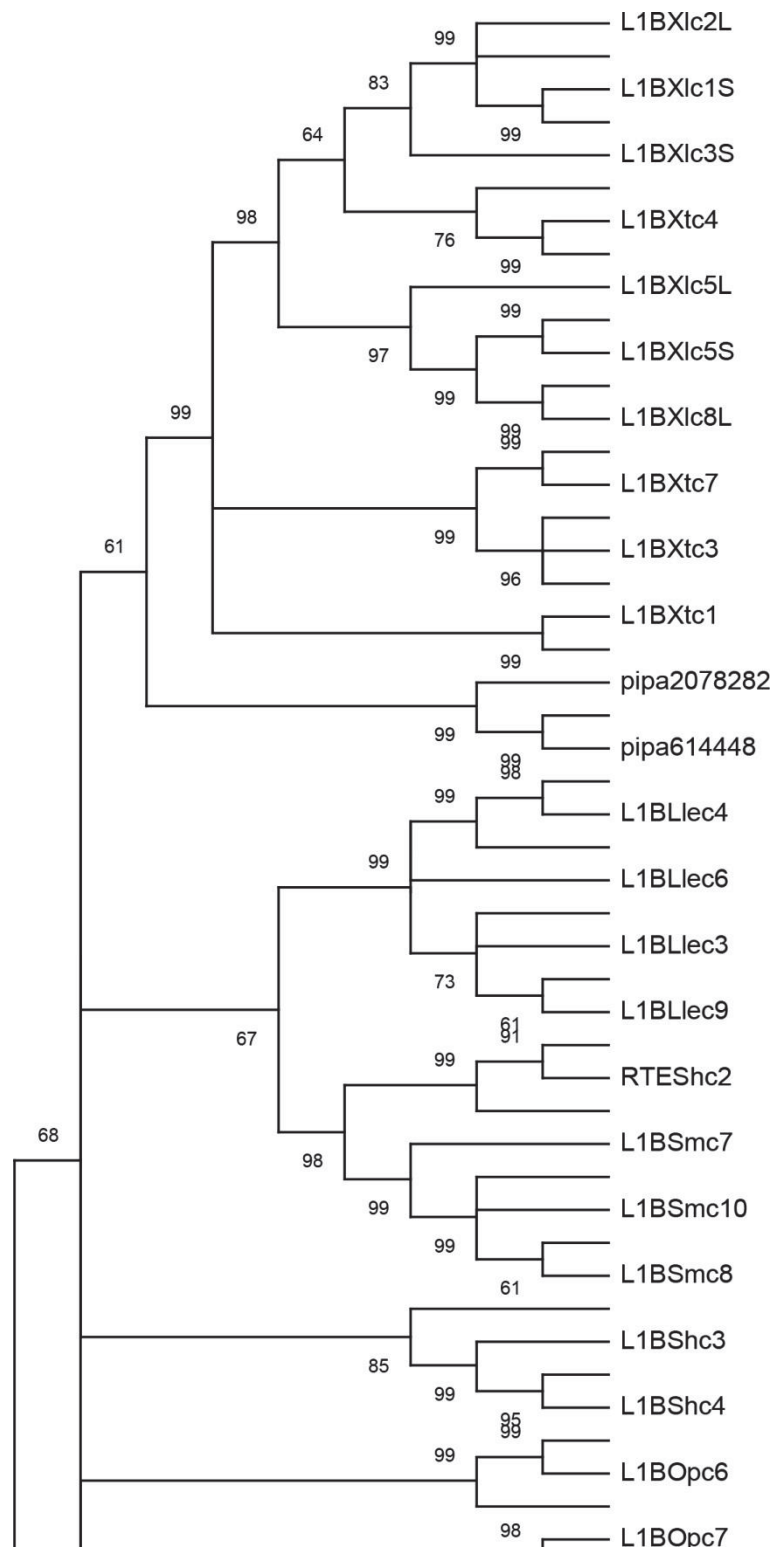
Anexo IV- Árvore filogenética com comprimento de ramo

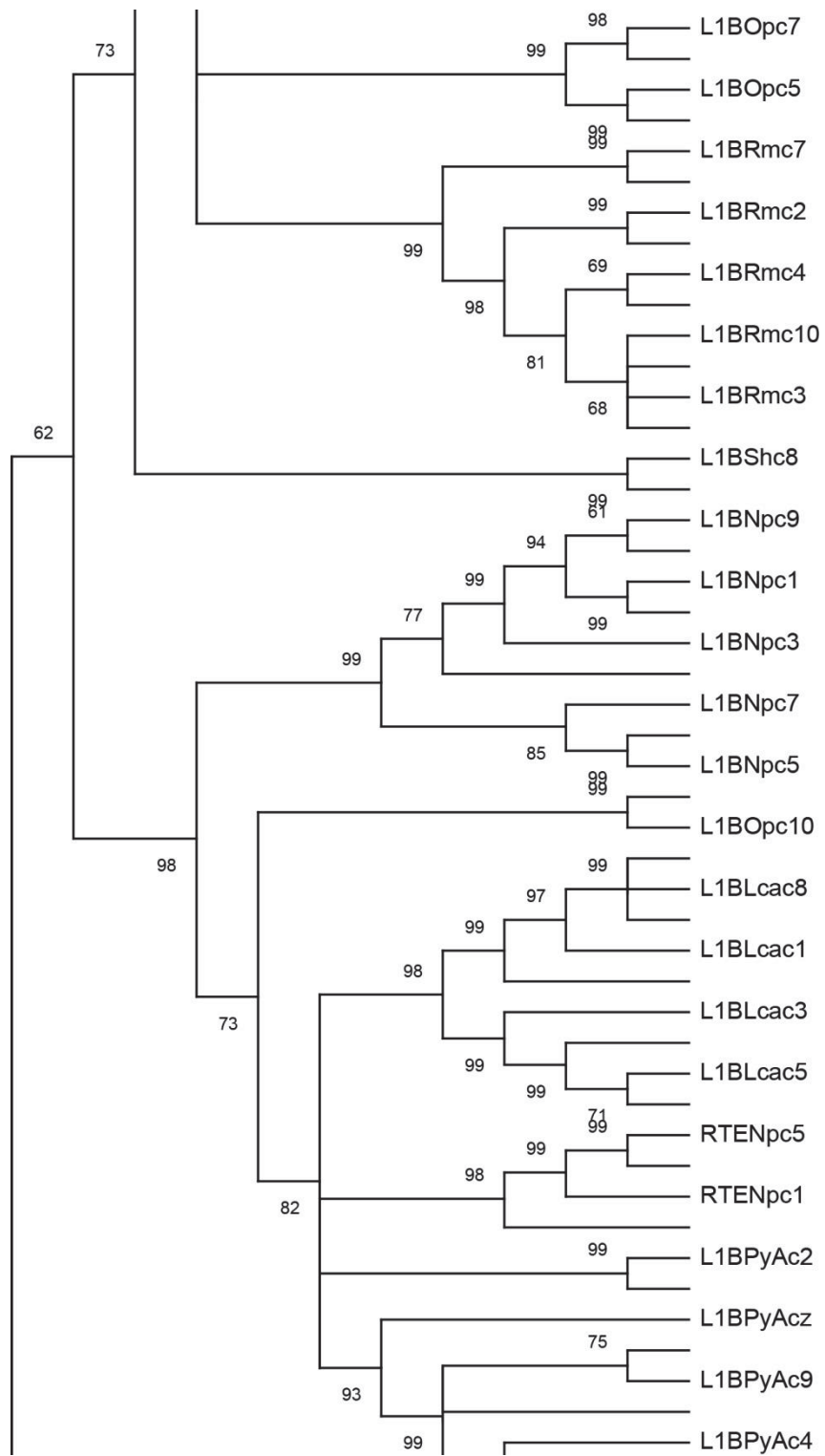
Árvore reconstruída por máxima verossimilhança (ML), modelo LG+G, a partir do domínio da RT dos elementos LINEs recuperados no genoma de *P. carvalhoi*. Os valores de *Bootstrap* são indicados próximo aos nós. Cada clado e espécie está representado por uma cor.

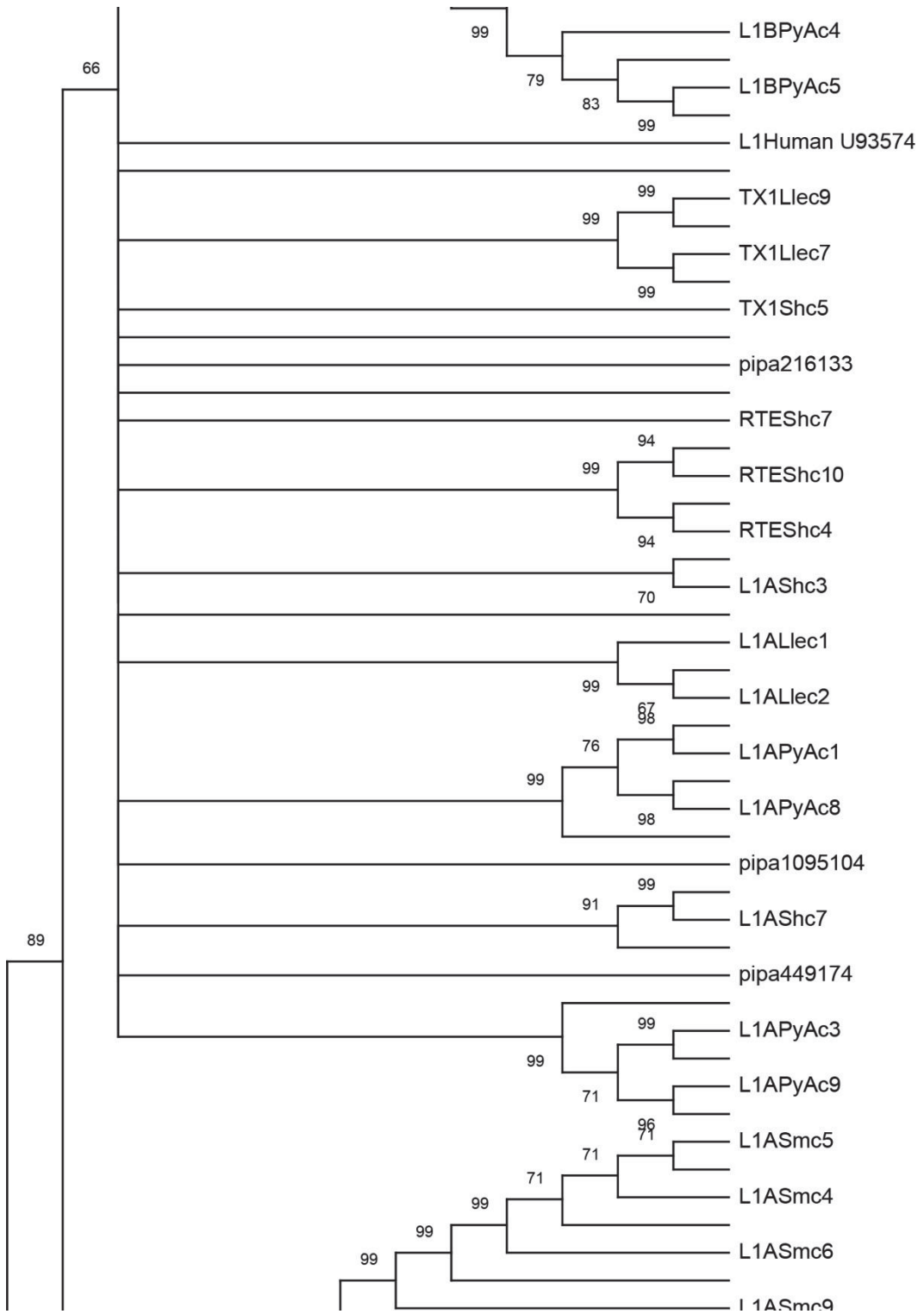


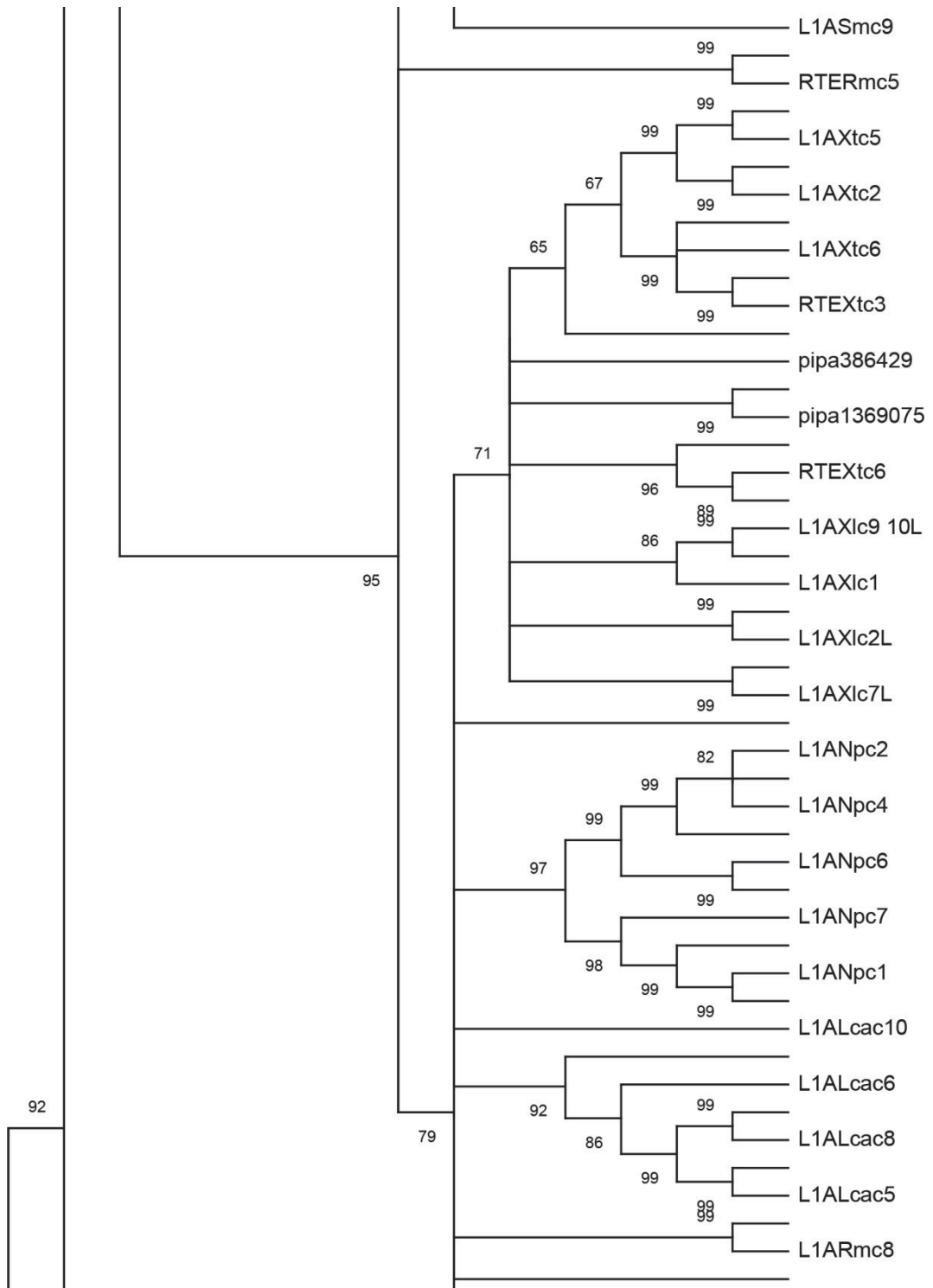
Anexo V Árvore preliminar exploratória inferida por *Neighbour Joining* (NJ)

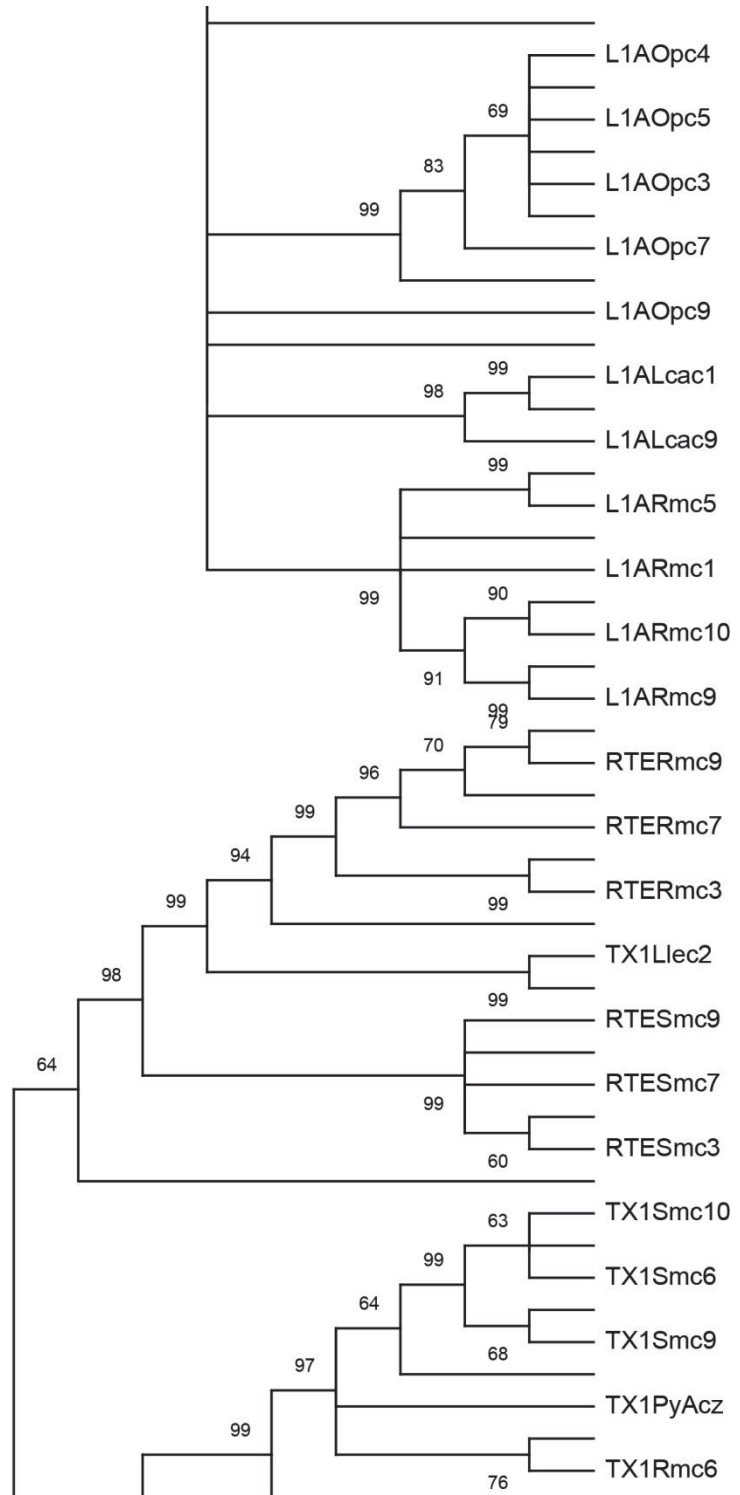
Construída utilizando modelo de *Poisson* (Anexo II), com 1000 réplicas de *bootstrap* foi obtida no programa MEGA X.

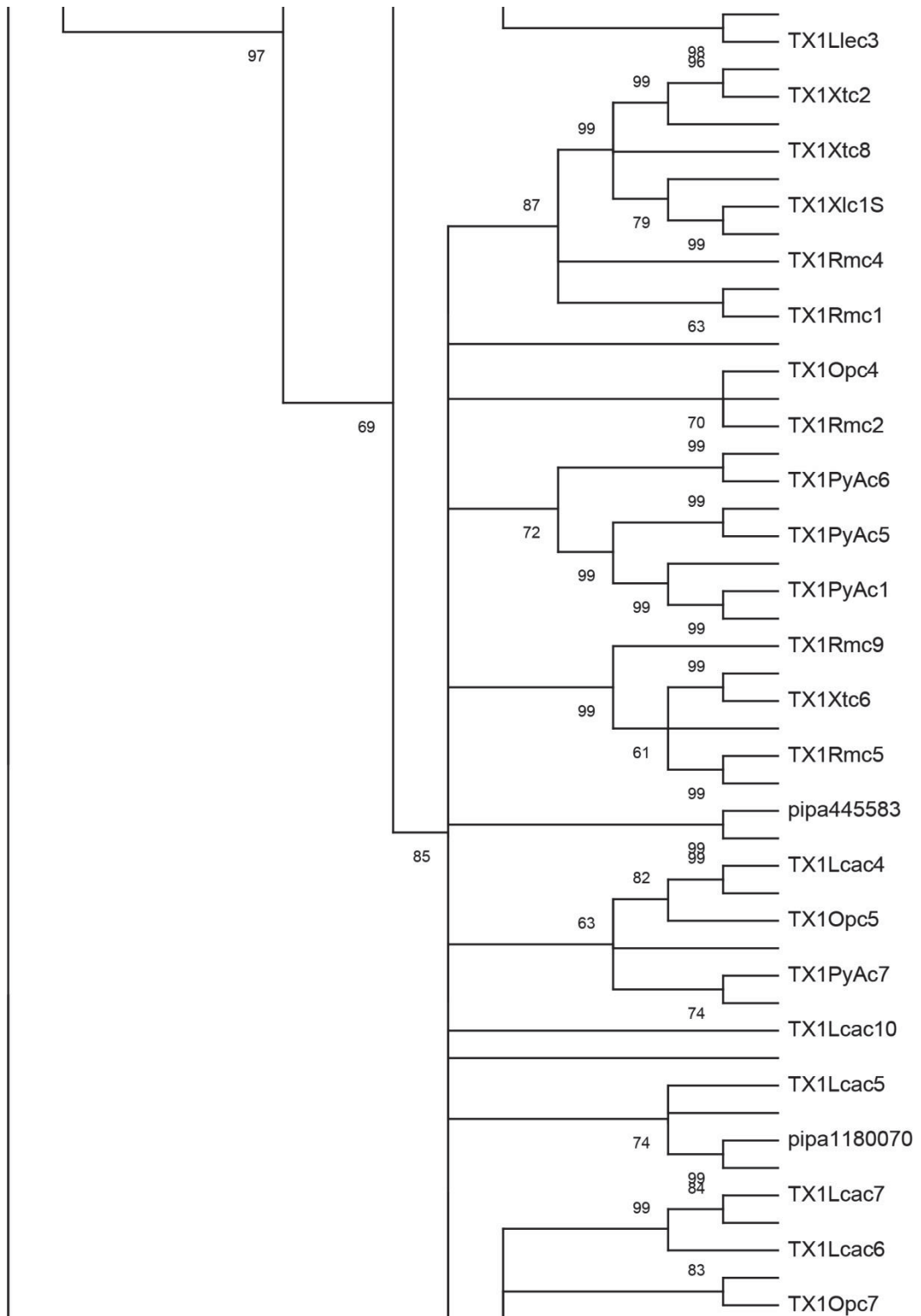


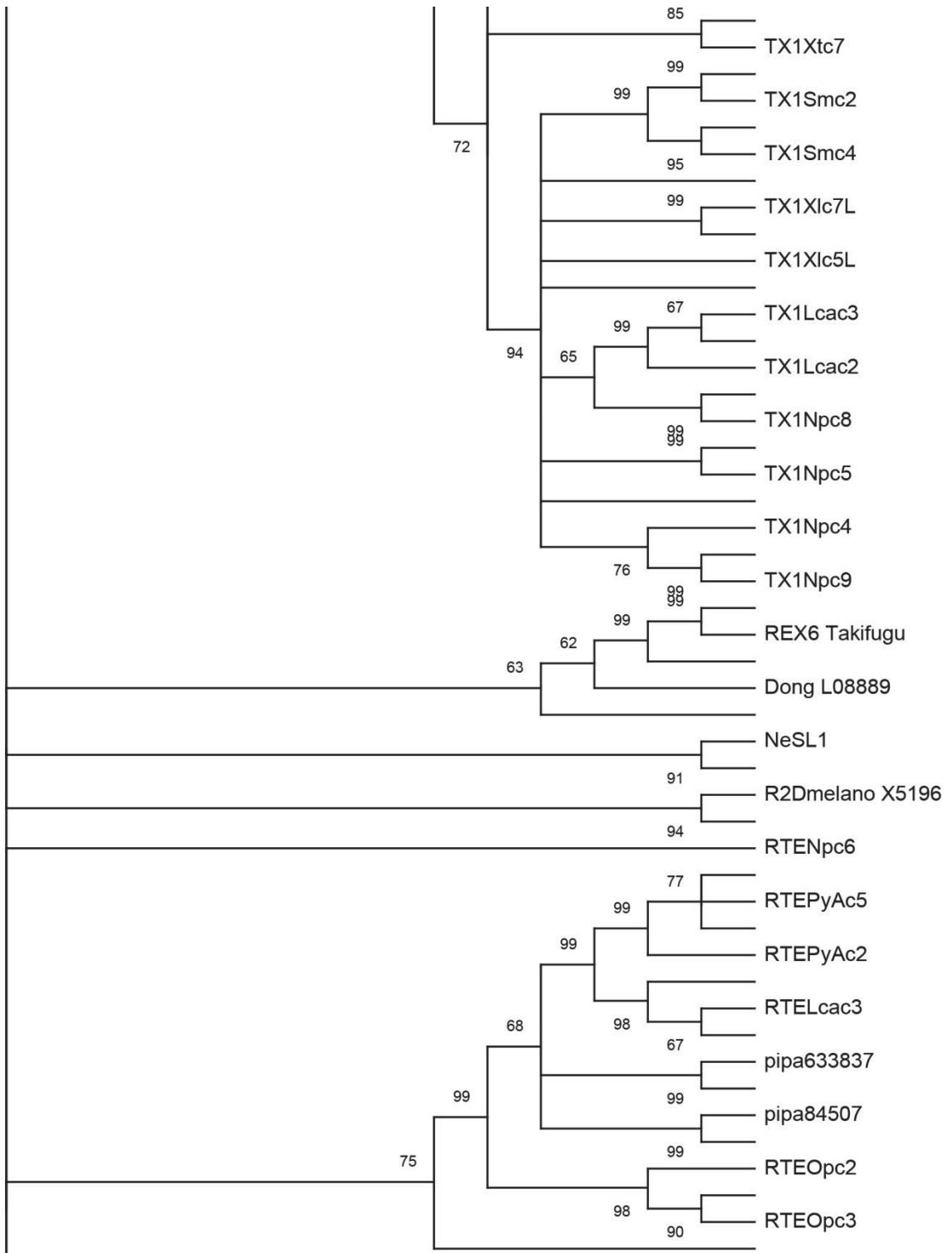


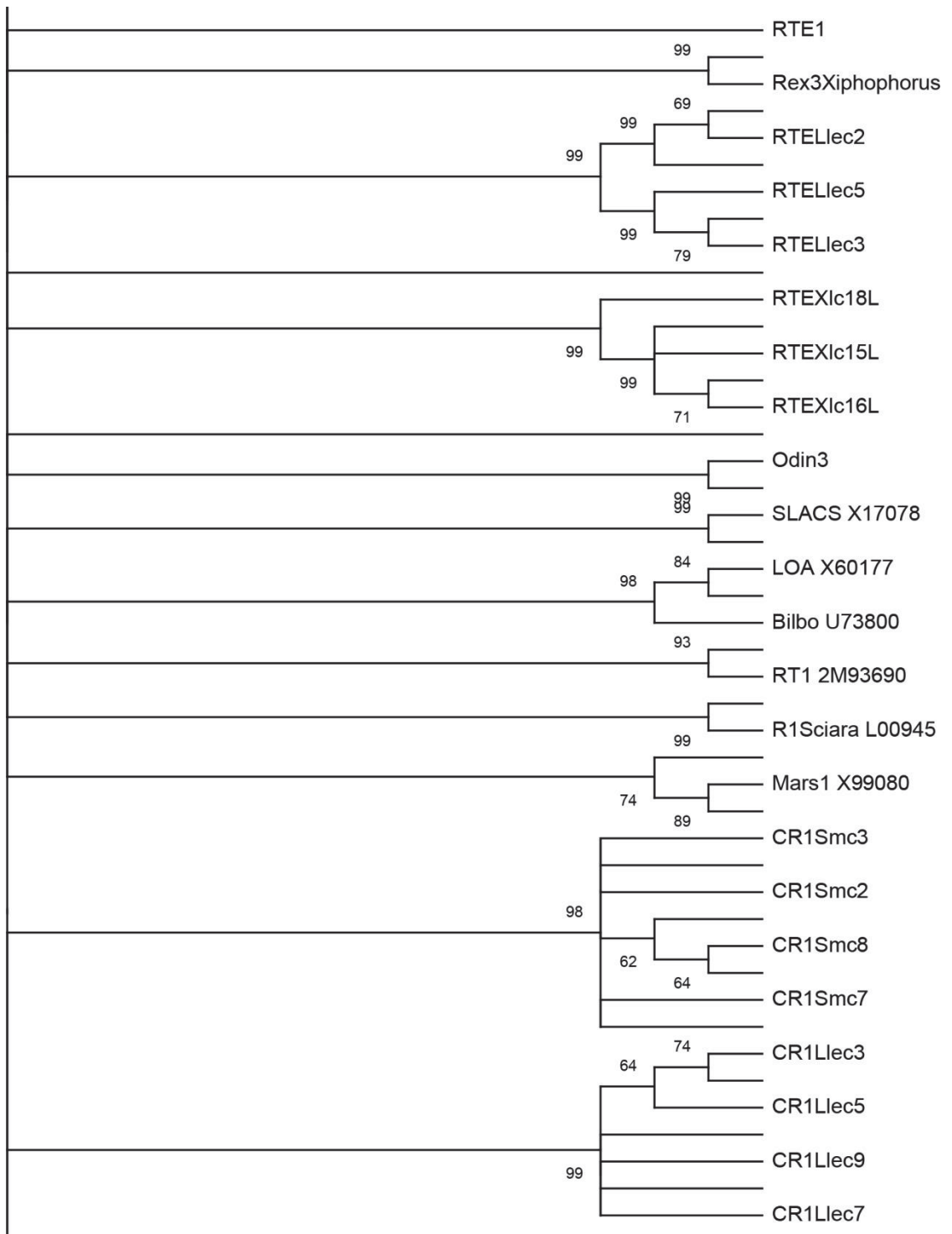


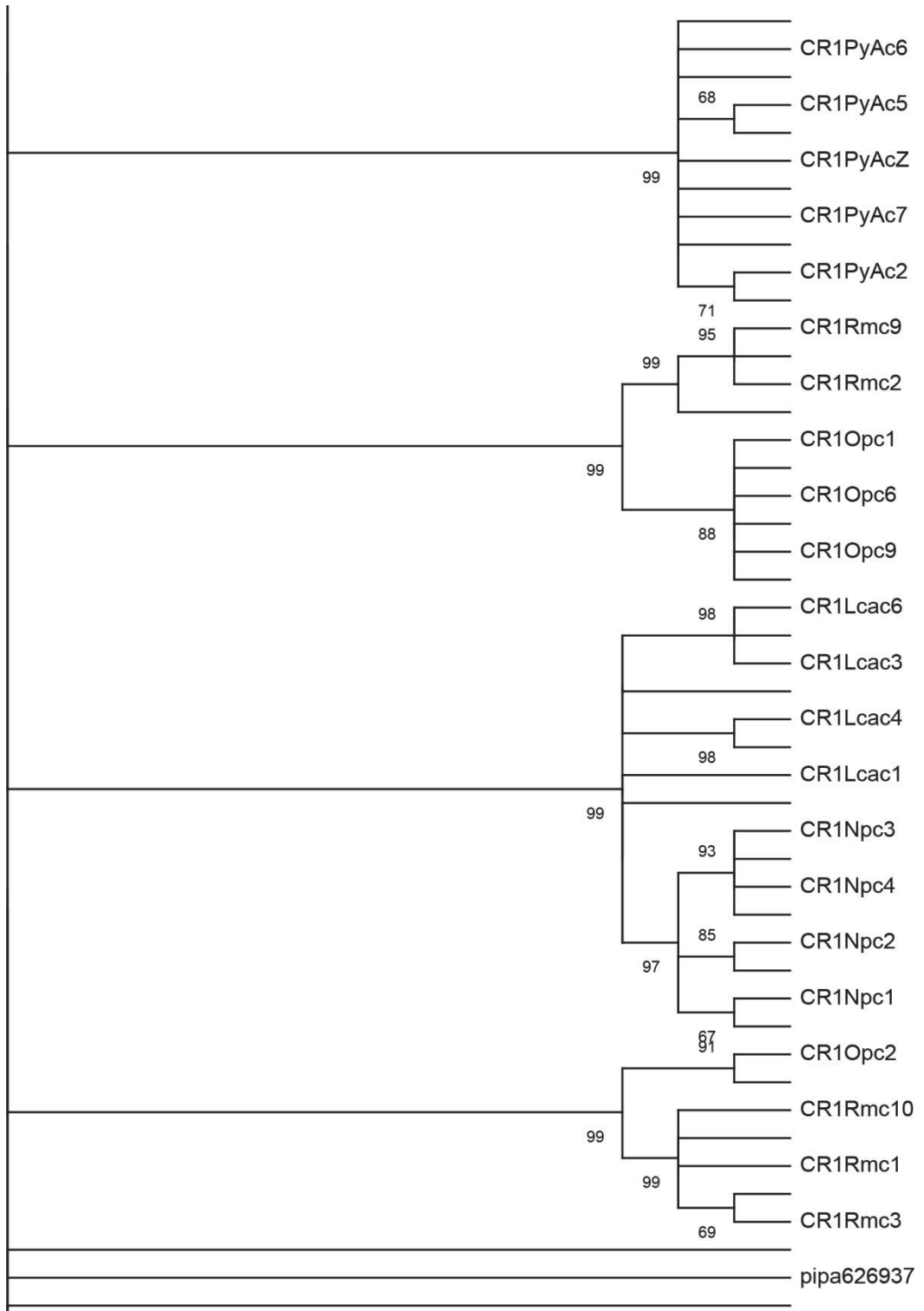


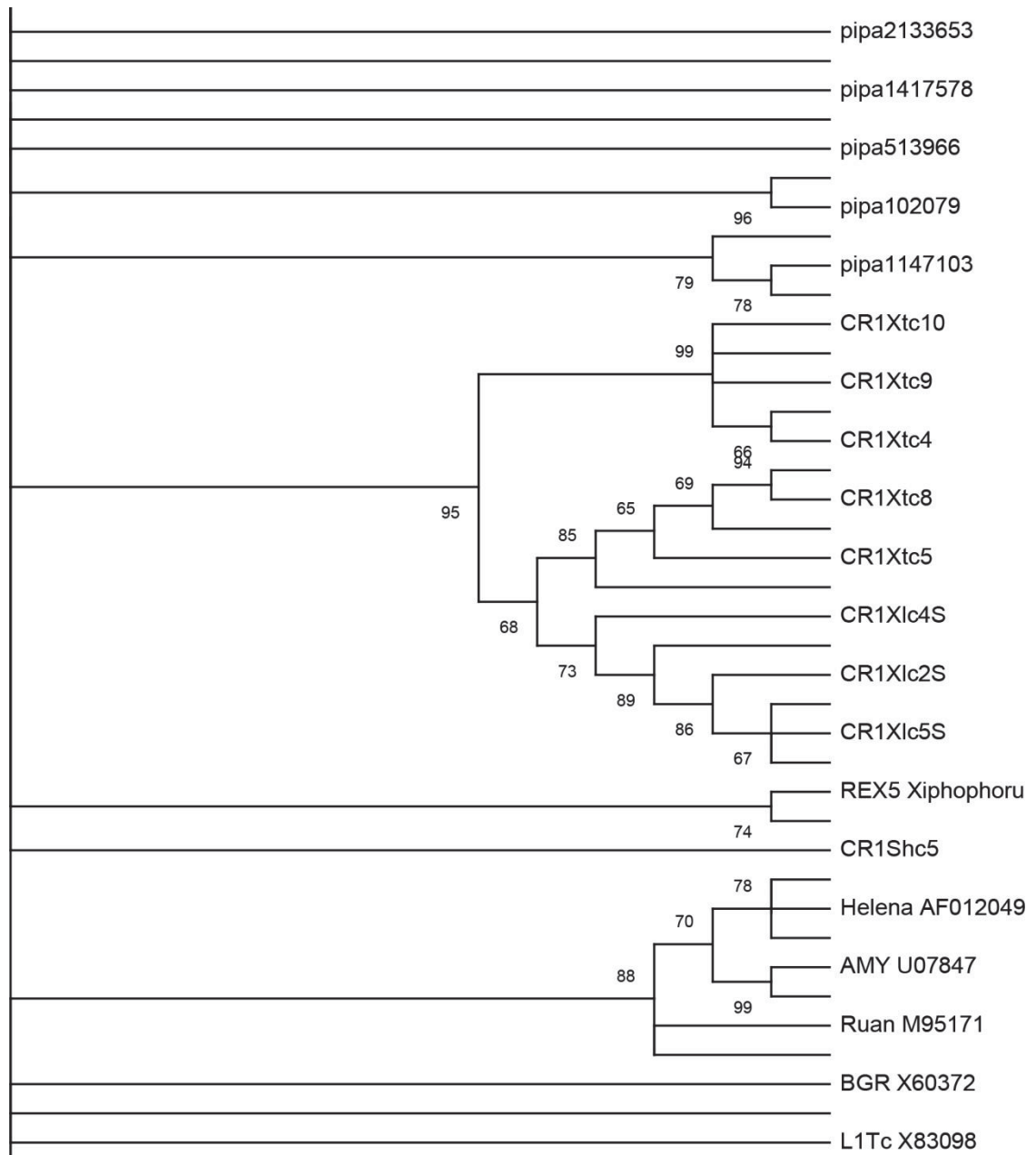


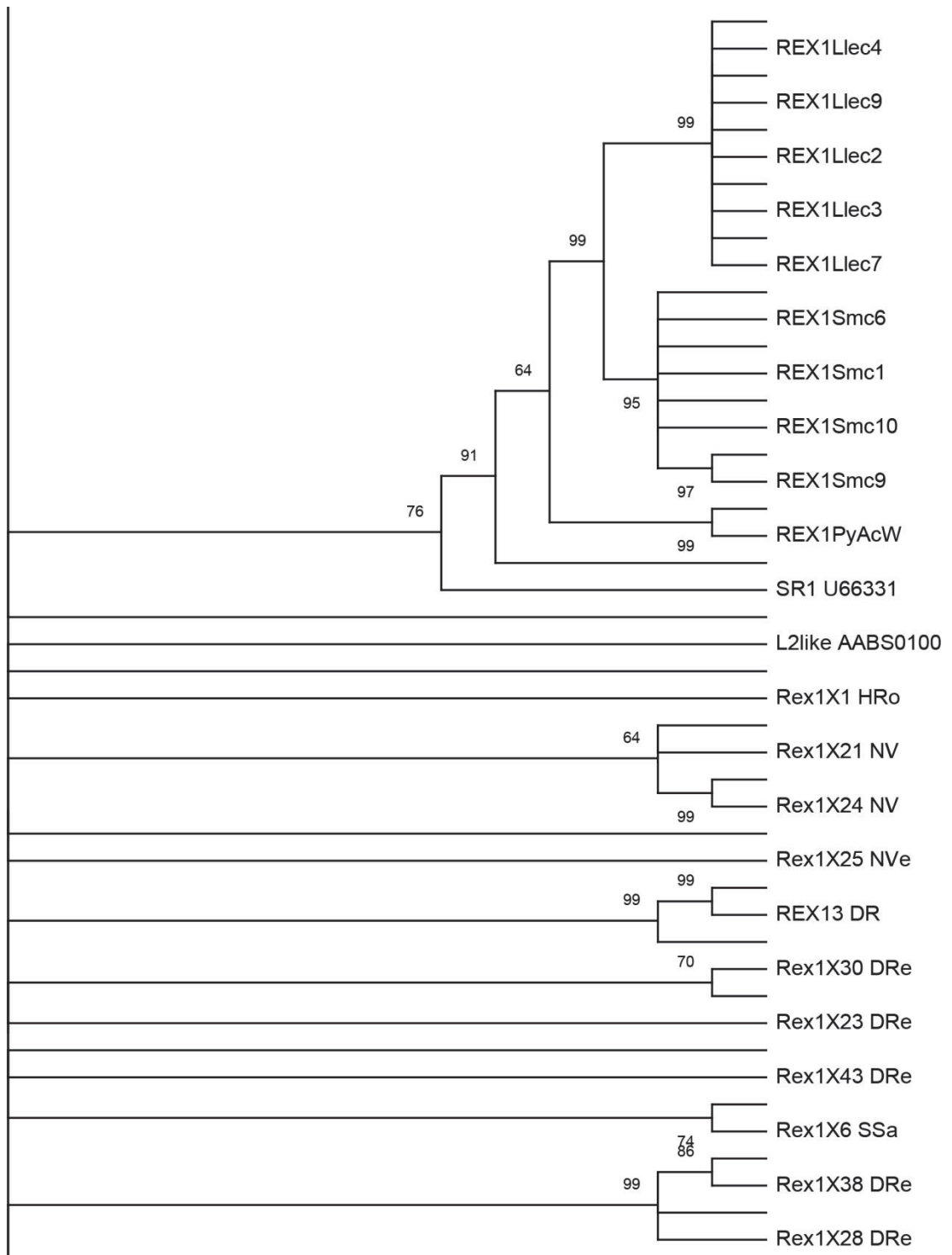


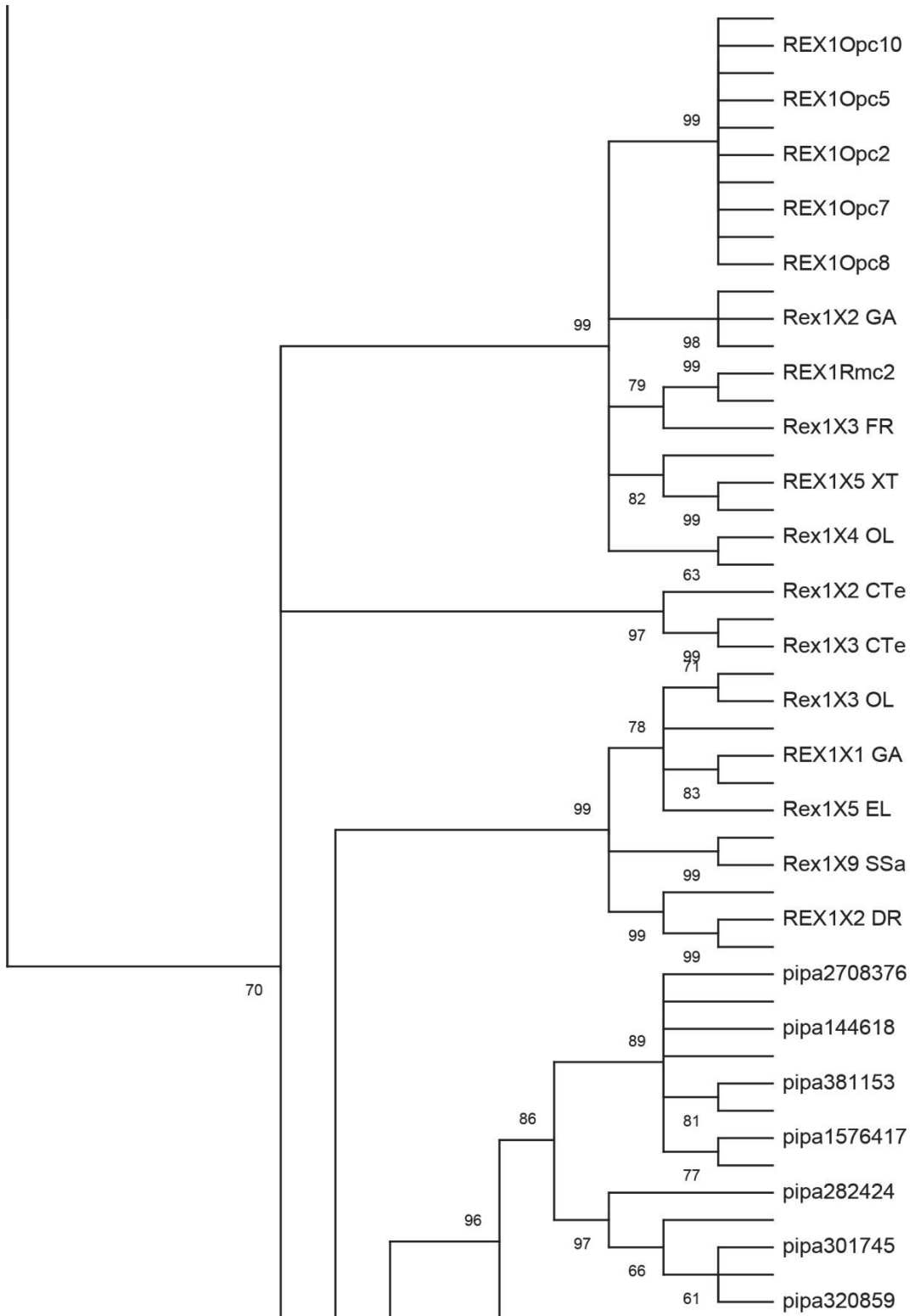


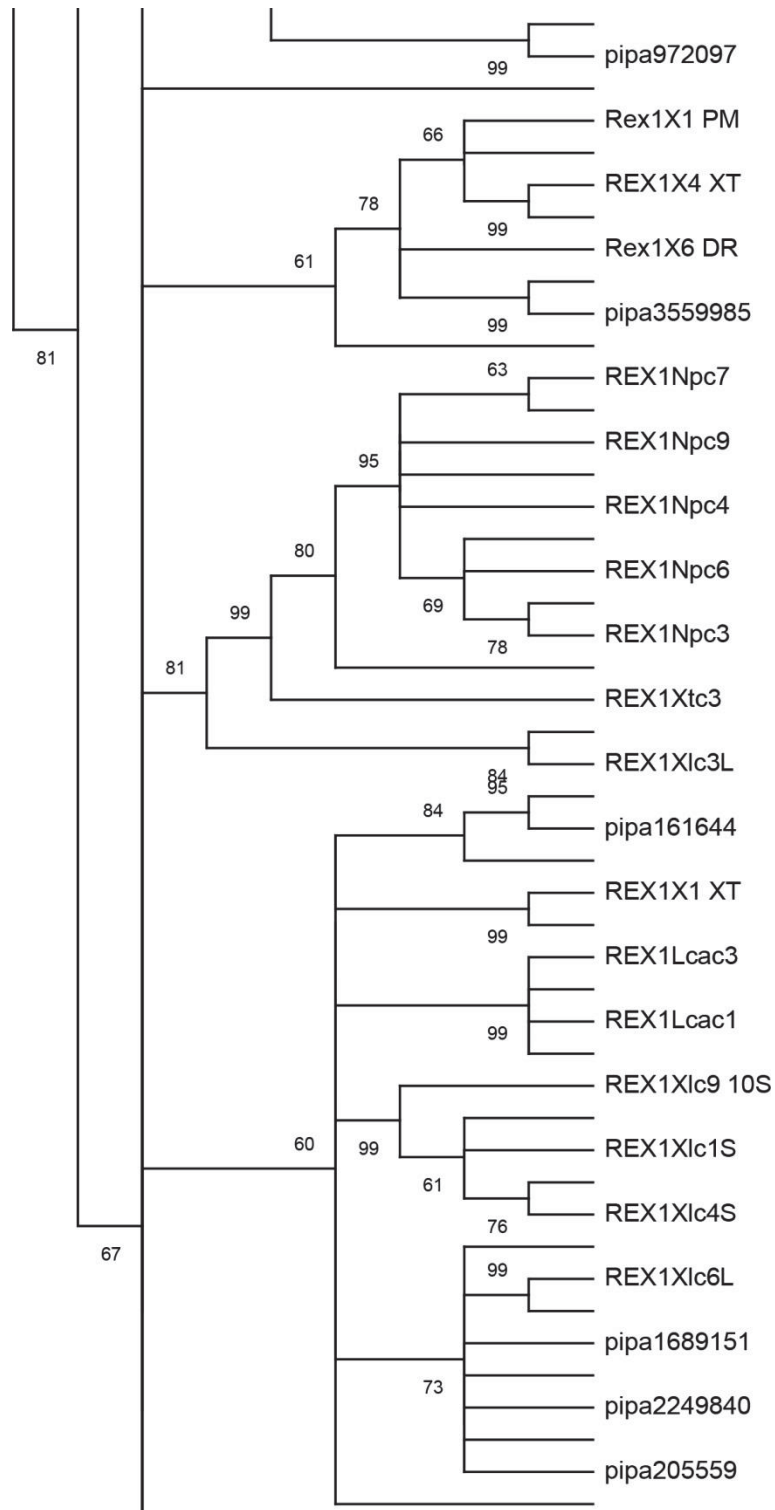


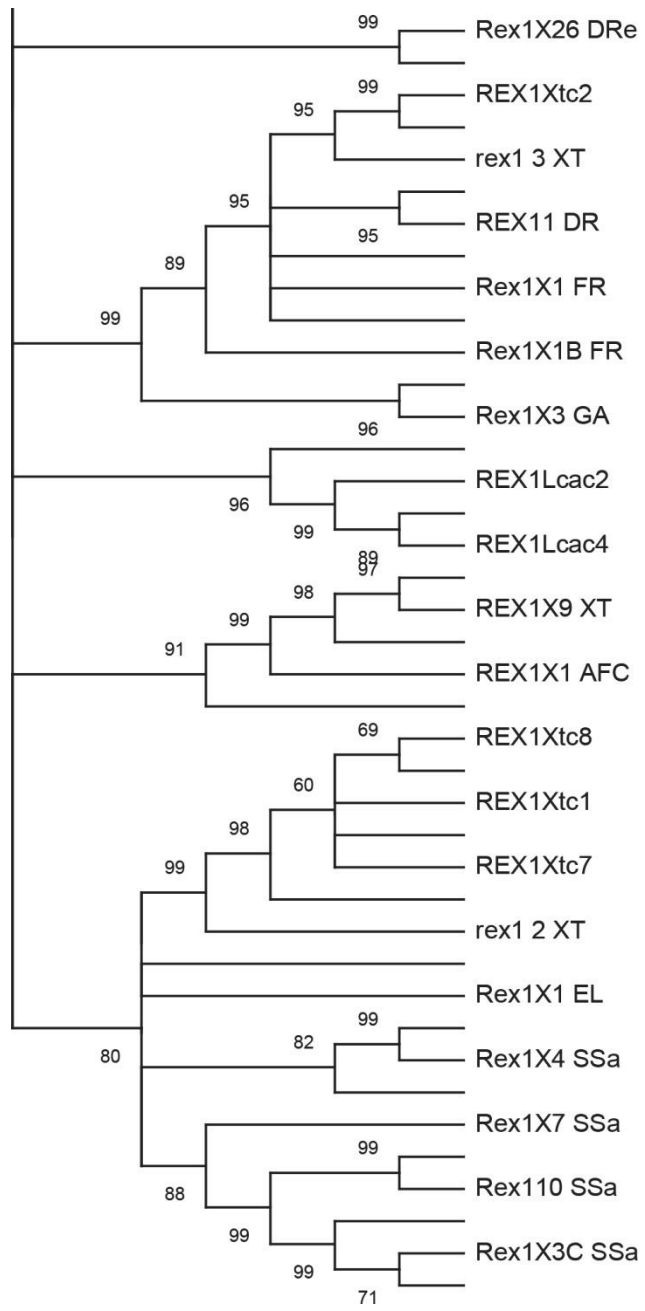












CAPÍTULO II

Isolamento, caracterização e mapeamento cromossômico de dois elementos da ordem LINE em *X. tropicalis* e *P. carvalhoi* (Anura, Pipidae)

Joana de Moura Gama¹, Adriana Ludwig², Camilla Borges Gazolla¹, Shirlei Maria Recco-Pimentel³ & Daniel Pacheco Bruschi¹.

1- Departamento de Genética, Universidade Federal do Paraná (UFPR); 2- Instituto Carlos Chagas; 3- departamento de Biologia Estrutural e Funcional, Universidade Estadual de Campinas (UNICAMP). Autor correspondente: danielpachecobruschi@gmail.com

Resumo: Grande parte do genoma de anuros estudados até o momento é composto por elementos de transposição, sendo predominantemente os elementos do tipo LTR e LINE, no entanto, existem poucos estudos sobre mapeamento físico dessas sequências em anuros. Nesse estudo, realizamos o mapeamento de dois elementos da ordem LINE (Rex1 e RTE) nos cariótipos de *P. carvalhoi* e *X. tropicalis*, duas espécies pertencentes à família Pipidae, através da Hibridização *in situ* Fluorescente (FISH). Nos cromossomos de *X. tropicalis* não foram identificados sinais de hibridação do elemento Rex1, já nos cromossomos de *P. carvalhoi* sinais de hibridações dessa sonda foram encontrados nas regiões pericentroméricas dos pares 3 e 7 e marcações subterminais nos pares 6 e 10. Ainda no par 6, sinais de marcações pontuais podem ser visto distribuídos ao longo da região intersticial do braço longo. Foram encontradas inserções em regiões de heterocromatina, como reportado anteriormente para elementos do tipo LINEs e associação com motivos de microssatélites, também comumente reportado para TEs. Já em *X. tropicalis*, possivelmente as cópias de Rex1 estejam presentes em menor densidade e dispersas entre os cromossomos, fato que justificaria a ausência de sinais de hibridação no cariótipo dessa espécie. Apesar de presentes no genoma de *P. carvalhoi*, nenhum sinal de hibridação do elemento RTE foi encontrado nessa espécie, fato que pode estar relacionado com o baixo número de cópias observado em nossas buscas genômicas. Já em *X. tropicalis*, nossos dados reforçam a sugestão de que nessa espécie o elemento RTE tenha sido perdido ou completamente degenerado durante a evolução. Nossos dados geram informações básicas sobre a distribuição e evolução desses elementos que podem contribuir para o entendimento do papel dessas sequências na evolução e diversificação desses genomas.

Palavras-chaves: LINE, Anura, Elementos de Transposição, Pipidae, Rex 1, RTE

INTRODUÇÃO

Anfíbios são conhecidos por apresentarem uma grande variação no tamanho dos genomas de suas espécies (GREGORY, 2020). Análises genômicas têm sugerido que grande parte dessa variação ocorre devido a proporção de elementos de transposição (TEs)

(LIEDTKE *et al.*, 2018; SUN *et al.*, 2015), ao elevado número de íntrons e/ou a complexidade de regiões regulatórias (LIEDTKE *et al.*, 2018).

Os TEs são sequências dispersas ao longo do genoma, sendo importantes para o processo de evolução dos mesmos (BIÉMONT; VIEIRA, 2006). A capacidade de mobilidade dos elementos pode levar a mudanças diretas na arquitetura do genoma, como produtos do processo de transposição quando esses elementos ainda se encontram ativos (BENNETZEN, 2005; DELPRAT *et al.*, 2009; LIM; SIMMONS, 1994) ou por meio de eventos de recombinação ectópica entre cópias altamente homólogas, podendo ocorrer nesse caso até mesmo após o elemento ter perdido sua capacidade de mobilização (FARRÉ *et al.*, 2011; GRAY, 2000; LARKIN *et al.*, 2006).

Grande parte do genoma de anuros estudados até o momento é composto por TEs, sendo predominantes os retrotransposons do tipo LTR e LINE (HELLSTEN *et al.*, 2010; SUN *et al.*, 2015). Na classe I (retroelementos) são encontrados os elementos da ordem LINEs (*Long Interspersed Element*) (WICKER *et al.*, 2007), que apresentam uma ORF codificando para uma Transcriptase Reversa (RT) e uma Endonuclease (EN). Além disso, em algumas superfamílias (Jockey, L1 e I), pode ser encontrada uma segunda ORF, posicionada anteriormente, comumente chamada de *gag* ou ORF1, tendo atividade de ligação a RNA. As regiões 5' e 3' não traduzidas (UTRs) dos LINEs são bastante variáveis, podendo haver ou não uma região promotora na região 5'UTR e podendo ter repetições curtas *in tandem* ou uma cauda poli (A) na extremidade 3' UTR (KAPITONOV; TEMPEL; JURKA, 2009; JURKA *et al.*, 2007; LÓPEZ-FLORES; GARRIDO-RAMOS, 2012). LINEs são amplamente distribuídos em vertebrados (CHALOPIN *et al.*, 2013, 2015; HAN, 2010; KAPITONOV; JURKA, 2008; VOLFF *et al.*, 1999, 2001), com uma diversidade desses elementos presentes no anuro *Xenopus tropicalis* (KAPITONOV; JURKA, 2008; GAMA *et al.*- Capítulo 1 em preparação), incluindo os elementos genericamente denominados Rex.

Além de *X. tropicalis*, elementos do tipo LINE estão presentes em várias espécies da ordem Anura, no entanto, estudos avaliando a localização cromossômica dessas cópias ainda são escassos. Nascimento e colaboradores (2015) isolaram e mapearam cópias do elemento Rex1 no cariótipo de *Physalaemus ephippifer* verificando uma organização clusterizada no cromossomo 3 dessa espécie. Já em espécies de peixes teleósteos, o mapeamento cromossômico de elementos LINEs (REX1, REX 3, e REX 6) tem sido realizado com maior frequência e tem se mostrado um excelente marcador para comparações cariotípicas

(BOUNEAU *et al.*, 2003; DASILVA *et al.*, 2002; FAVARATO *et al.*, 2017; FISCHER *et al.*, 2004; OZOUF-COSTAZ *et al.*, 2004).

Nesse trabalho, avaliamos a localização cromossômica de cópias de dois elementos da ordem LINE (Rex1 e RTE) no cariótipo de duas espécies de anuros da família Pipidae. O anuro *Xenopus tropicalis*, sendo considerado grupo irmão do gênero *Pipa* (IRISARRI *et al.*, 2011), é comumente utilizado em estudos genômicos e genéticos, por se tratar de modelo animal (GRAINGER, 2012). Desse modo, possui genoma e cariótipo bem estudados e compartilha do mesmo número diploide ($2n=20$) com *Pipa carvalhoi*, um anuro sul-americano, endêmico da Mata Atlântica, de habitats aquático e objeto principal do nosso estudo (CANNATELLA; TRUEB, 1988; SILVA *et al.*, 2010, FROST, 2020).

MATERIAL E MÉTODOS

Caracterização das cópias de Rex1 e RTE

Para a caracterização molecular dos retroelementos Rex1 e RTE, foram utilizados os *contigs* recuperados em GAMA, JM(em preparação - Capítulo 1), a partir de um montagem preliminar do genoma de *P. carvalhoi*. As buscas por essas sequências foram realizadas através da ferramenta tBLASTn local (ALTSCHUL *et al.*, 1997) no genoma de *P. carvalhoi*. Foram analisados os 15 *contigs* com maior significância (maior score e *e-value* menor que 10^{-4}). A presença de ORFs foi analisada pela ferramenta do NCBI, ORFinder (<https://www.ncbi.nlm.nih.gov/orffinder/>), e os domínios proteicos foram preditos através da ferramenta CD-Search do NCBI (MARCHLER-BAUER; BRYANT, 2004). Os elementos foram considerados completos quando possuíam domínios de uma EN e RT que não estavam truncados ou degenerados.

Construção dos primers e preparação das sondas de Rex 1 e RTE

Para a construção dos *primers* foram utilizadas sequências da RT dos elementos Rex 1 e RTE, isoladas a partir da versão rascunho do genoma de *P. carvalhoi* isoladas no artigo Gama *et al.*, (em preparação -Capítulo 1). As diferentes cópias de cada elemento foram alinhados na ferramenta MUSCLE *online* (<https://www.ebi.ac.uk/Tools/msa/muscle/>) (MADEIRA *et al.*, 2019) e as sequências dos *primers* foram selecionadas manualmente no programa GeneDoc 2.7 (NICHOLAS; NICHOLAS; DEERFIELD, 1997), considerando sempre regiões mais as conservadas entre as cópias. A verificação de parâmetros como

temperatura de anelamento, formação de grampos e complementaridade foi realizada utilizando *OligoCal 3.27* (KIBBE, 2007). As sequências dos *primers* são apresentadas na Tabela 1.

Tabela 1. Sequência dos primers utilizados nesse trabalho com informações dos *contigs* utilizados em sua construção, a temperatura de anelamento (°C) e o tamanho dos fragmentos estimados em pares de bases (pb).

Elemento	Contig	Sequência	Temp. anelamento	Tamanho
Rex1	pipa626994/pipa128038	F: TCA RGA AGC TAA ATG CAC TGG	57.5-59.5	760
		R: A CCC ACA ATG CTT AGT GC	58.4°	
RTE	pipa84507/pipa1532542	F: ATC CTG GAG CAG ATT GCG	56.3	524
		R: GTC CTG GTC TTT GAT GGC	56.3	

Para o isolamento das sequências do genoma, o DNA total de *P. carvalhoi* foi extraído utilizando o método TNES (BRUSCHI *et al.*, 2012). As amostras foram amplificadas através de uma reação em cadeia da polimerase (PCR) com volume total de 25 µl, *Buffer* 1x, 3 mM de MgCl²⁺, 0,28 pmol de *primer forward* e 0,28 pmol de *primer reverse*, 32U de Taq DNA polimerase (500U), 25 ng/µl de DNA e 0,5 ng/µl de DNTP. Após testes de gradientes de temperatura (54-60°C) para verificação da melhor temperatura de anelamento, a seguinte condição foi utilizada: 94°C por 5 min seguido de 35 ciclos de 94°C por 1 min, 60°C por 50 seg, e 72 °C por 1 min, seguido por uma extensão final de 72°C por 8 min.

O produto de PCR foi utilizado como *template* para a preparação das sondas de Rex1 e RTE. As sondas foram marcadas por PCR através da incorporação de dUTP diretamente marcados com 5-TAMRA-dUTP de acordo com protocolo descrito por Guerra (2012).

Preparações cromossômicas e experimentos de FISH

Os espécimes de *P. carvalhoi* analisados foram coletados no Flona do Apodi, na cidade de Cratos, Ceará, Brasil (7°21'55.80"S, 39°26'26.66"W), sob licença do SISBIO/Instituto Chico Mendes de Conservação da Biodiversidade (protocolo 55481-1) e os espécimes foram depositados no Museu de História Natural da Universidade Federal do Alagoas (MHN-UFAL).

Para a obtenção das preparações cromossômicas de *P. carvalhoi*, os anuros foram submetidos a uma injeção intraperitoneal de colchicina 2% (proporção de 0,02 mL/g do

animal), 4 horas depois as suspensões de intestino e testículos foram obtidas de acordo com os protocolos de King and Rofe (1976) e Schmid (1978), adaptados.

As amostras de suspensões celulares de *X. tropicalis* foram obtidas através de cultura celular de fibroblastos desenvolvidas no laboratório do Dr. Malcolm A. Ferguson-Smith, do *Department of Veterinary Medicine da University of Cambridge* e cedidas pelo pesquisador Dr. Nicolas Pollet, do *Evolution Genomes Comportement Ecologie Lab*, da Paris South University (UPS).

Os experimentos de hibridação *in situ* fluorescente foram conduzidos utilizando o protocolo de Viana *et al.* (2017) com modificações. As lâminas contendo as metáfases foram submetidas a solução de PBS 1X por 3 min em temperatura ambiente e posteriormente em série alcoólica (70%-80%-absoluto) por 2 min. Após foram adicionadas em RNase 0,4% a 37°C por 1 hora e lavadas 3 vezes em SSC 2X por 3min, seguido de uma lavagem em PBS 1X por 3 min. A fixação foi realizada com formaldeído 1% em temperatura ambiente por 10 min, seguido de novo banho em PBS 1X por 3 min e novamente submetidas a série alcoólica nas mesmas condições já citadas. A desnaturação do DNA foi realizada em formamida 70% por 3 min, seguida de série alcoólica gelada por 2 min.

A lâminas foram deixadas hibridando por 72h em câmara escura umedecida com formamida, com uma concentração de 40 ng/ul de sonda. Após as 72h de hibridação as lâminas foram lavadas em formamida 15% a 42°C por 7 min, seguidas de 2 banhos em tween 0,5% por 3 min em temperatura ambiente e novamente submetidas a série alcoólica de 2 minutos. A detecção foi realizada de forma direta utilizando 5-TAMRA-dUTP seguido de contracoloração por DAPI. Todos os experimentos foram realizados em duplicatas e sempre utilizando em um mesmo experimento metáfases de ambas as espécies e as lâminas foram analisadas e fotografadas com o auxílio do Microscópio Olympus.

RESULTADOS E DISCUSSÃO

Caracterização molecular das cópias de Rex1 e RTE

Isolamos os *contigs* referentes aos elementos recuperados para o *cluster* de Rex1 na árvore de sequências recuperada por Gama, JM (em preparação – Capítulo 1). Cada sequência recuperada foi analisada individualmente em buscas de assinaturas moleculares para sua caracterização (Anexo III- Capítulo 1). As sequências provenientes do agrupamento com

Rex1 apresentam grande divergência em nível de nucleotídeo quando comparadas entre si e agrupam todas no mesmo grande *cluster* (Rex1) (GAMA, JM- em preparação capítulo 1). Na maior parte das cópias, apenas o domínio RT foi encontrado, sendo que apenas um *contig* (denominado *pipa381153*) (Figura 1 A) apresentou os domínios da EN e RT, porém, em ORFs separadas, enquanto no elemento completo/ativo esses dois domínios deveria estar na mesma ORF (KAPITONOV; TEMPEL; JURKA, 2009). Também, não foi possível o estabelecimento dos limites desse elemento. A cópia mais completa perfaz todo o *contig* e não há sequência suficiente para evidenciar a presença de uma ORF adicional na região 5’.

A ausência da região da EN nas demais cópias pode indicar que as cópias estão truncadas, o que é uma característica de LINEs já degenerados no genoma. Isso acontece devido seu modo de inserção TPRT (*Target-primed reverse transcription*) onde é comum a interrupção prematura da transcrição, gerando essas extremidades truncadas (JURKA *et al.*, 2007; LÓPEZ-FLORES; GARRIDO-RAMOS, 2012), justificando o porquê de encontrarmos apenas um *contig* com domínio da EN. Esse fato também já foi relatado por Volff *et al.* (2000), que encontrou esse domínio apenas para os elementos Rex1-*Ten*, identificado no genoma do peixe teleósteo *Tetraodon nigroviridis*, enquanto nos outros elementos Rex1 do genoma de diversos peixes teleósteo, apenas o domínio da RT fora encontrado, assim como em *P. carvalhoi*. Além disso, não foi possível recuperar repetições do tipo microssatélites na extremidade 3’, característica apontada para Rex-1 de peixes e de *X. tropicalis* (Figura 1-B).

Segundo informações disponíveis no Repbase (BAO; KOJIMA; KOHANY, 2015) elementos Rex1 de *X. tropicalis*, por exemplo, podem ser subdivididos em 9 tipos (Rex1-1 a Rex1-9) e variam na sua conservação e na composição de microssatélites presentes na extremidade 3’ UTR. Possivelmente, estamos trabalhando apenas com cópias incompletas devido ao alto nível de fragmentação do genoma, pois o maior *contig* contendo sequências de Rex1 é de 3305 pb e a ORF da RT se expande quase até o final desse *contig* (Figura 1A). Dessa forma, é possível que a ausência das regiões terminais seja devida a montagem, que pode ter sido prejudicada pela própria natureza da sequência repetitiva.

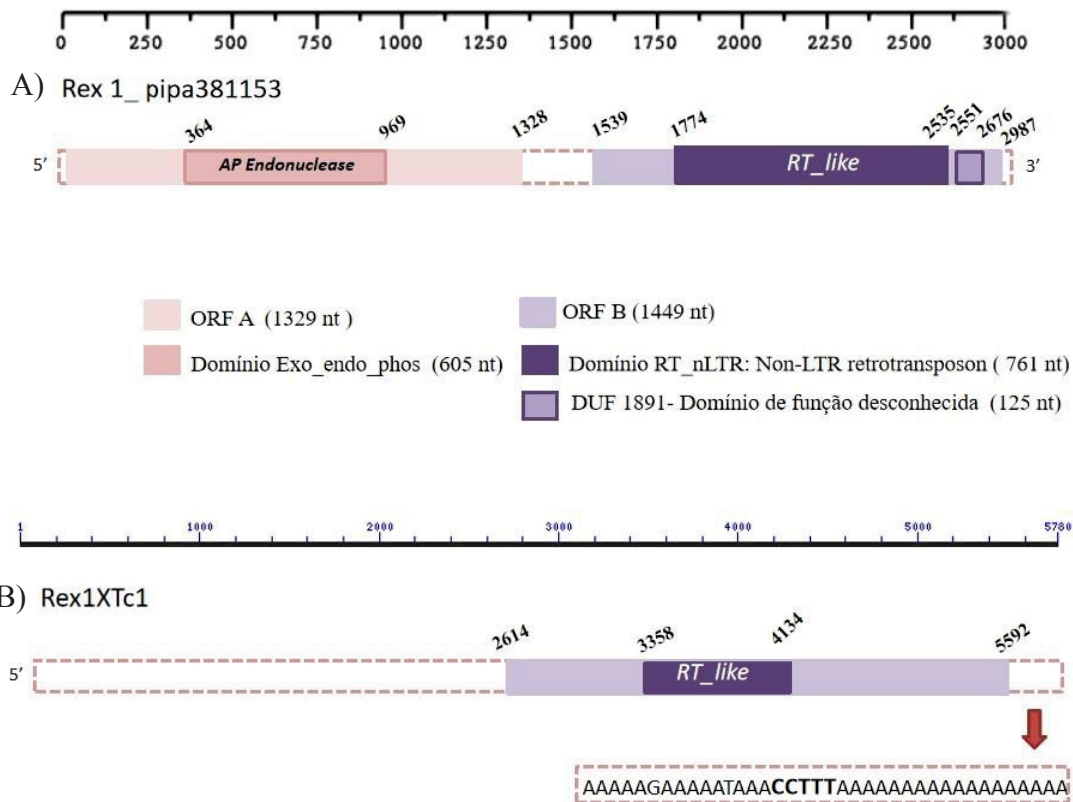


Figura 2. Representação de regiões genômicas contendo cópias do do Retroelemento Rex1. A) Rex1 de *P. carvalhoi* desenhado com base no *contig* pipa381153. B) Rex1 de *X. tropicalis* desenhado com base na sequência presente no cromossomo 5. As linhas pontilhadas em rosa representam a região cromossômicas, as ORFs estão representadas de uma mesma cor (conforme legenda) e seus domínios de cores mais escuras. A régua acima ilustra o tamanho do *contig*/região cromossômica isolada, evidenciando as posições das ORFs e domínios. Na figura 1B está destacado o final do elemento contendo a cauda Poly-A.

Em relação ao elemento RTE, recuperamos apenas três *contigs* (*pipa633837*, *pipa2448609* e *pipa84507*) na porção do genoma analisada, todos eles apresentaram o domínio da RT e apenas um deles não possuía o domínio completo (*pipa 2448609*). O único *contig* que apresentou os dois domínios característicos dos LINES (EN e RT) foi o *pipa84507*, que curiosamente foi a cópia agrupada em nossa filogenia com RTE proveniente da

sanguessuga *Helobdella robusta* (GAMA *et al* - em preparação – Capítulo I). O mesmo *contig* foi utilizado para o desenho representativo do elemento RTE de *P. carvalhoi* (Figura 2) e nenhuma repetição foi encontrada no final do elemento, como reportado para elemento Rex3 de peixes (pertencente também ao clado RTE) (VOLFF *et al.*, 1999). No entanto, da mesma forma que para Rex1, a ausência dessas repetições não significa que elas não estejam presentes, podendo decorrer do estado de montagem do genoma em estudo. Esse problema só será resolvido com uma melhor montagem do genoma desse anuro. Apesar de nossas estratégias de busca por tBlastn e análise filogenética (GAMA *et al*, em preparação - Capítulo 1) nenhuma cópia de elemento RTE foi encontrada no genoma de *X. tropicalis*, reforçando os resultados apontados por Sun *et al.* (2015).

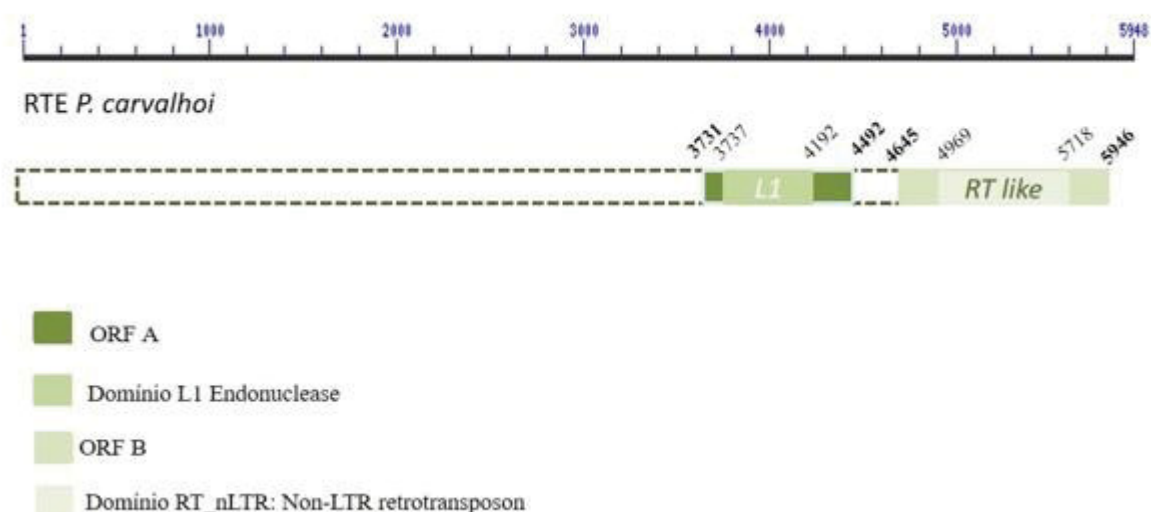


Figura 3. Representação esquemática da região genômica contendo cópia do LINE RTE de *P. carvalhoi*, baseado no *contig* pipa84507. As linhas pontilhadas (em verde escuro) representam a região cromossômica, as ORFs estão representadas de uma mesma cor (conforme a legenda) e seus domínios de cores mais escuras. A régua acima ilustra o tamanho do *contig* isolado evidenciando as posições das ORFs e domínios.

Mapeamento cromossômico das cópias de Rex1 e RTE

Experimentos de FISH utilizando a sonda dos elementos Rex1 nos cromossomos de *P. carvalhoi* revelou a presença de sinais de hibridação nos homólogos dos pares 3, 6, 7 e 10 (Figura 3A). Nos pares 3 e 7, as marcações se limitam aos centrômeros, enquanto nos pares 6 e 10 as marcações estão nas regiões subterminais dos cromossomos. Ainda no par 6, podemos ver sinais de marcações pontuais distribuídos ao longo da região intersticial do braço longo desses homólogos (Figura 3A). Apesar de detectarmos sinais de hibridação em outros cromossomos do cariótipo (por exemplo, cromossomos 1, 2 e 4), essas marcações estavam

limitadas a apenas um dos cromossomos do par. Já no cariótipo de *X. tropicalis*, nossos experimentos de FISH não revelaram a presença de sinais de hibridação em nenhum dos cromossomos desse cariótipo (Figura 3B). Para nos certificarmos de que a ausência desses sinais não representaria um artefato metodológico, todos os experimentos de FISH foram realizados sempre em um mesmo dia com ambas as espécies e mesma sonda.

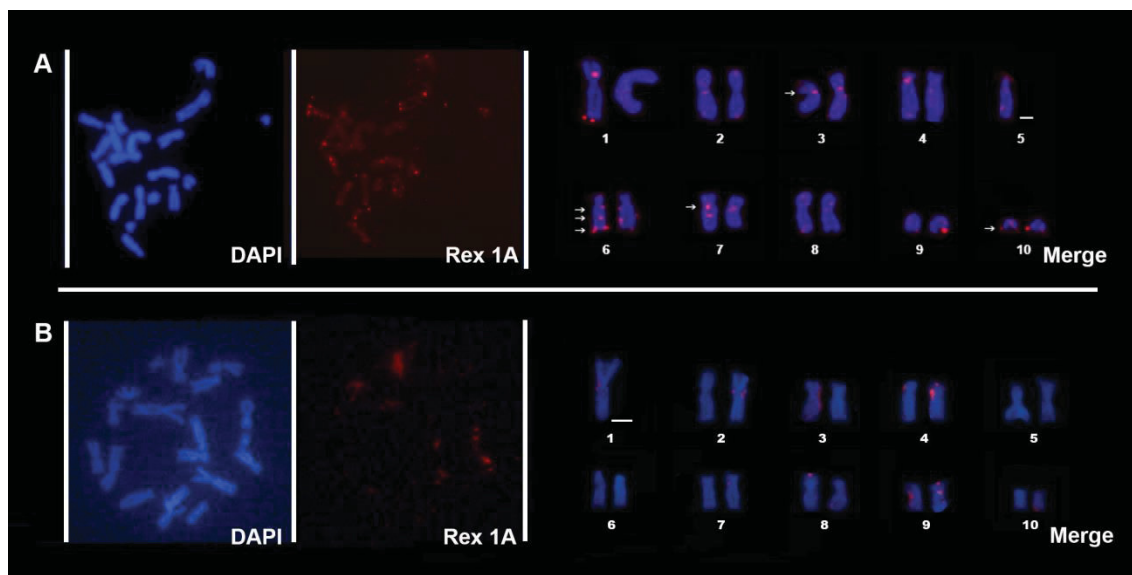


Figura 4: Mapeamento dos elementos Rex 1 nos cromossomos de *P. carvalhoi*. (A) e *X. tropicalis* (B). As setas evidenciam os sinais considerados como marcação do elemento na figura A. Em ambos casos, metáfase parcial é apresentada.

Os elementos Rex1 são amplamente estudados e mapeados nos cromossomos de peixes (FAVARATO *et al.*, 2017; OZOUF-COSTAZ *et al.*, 2004; TRALDI *et al.*, 2019; VALENTE *et al.*, 2011), estando preferencialmente presentes em regiões ricas em heterocromatina, como nas porções pericentroméricas e subterminais dos cromossomos (OZOUF-COSTAZ *et al.*, 2004; VALENTE *et al.*, 2011). No cariótipo de *P. carvalhoi* podemos notar esse padrão, onde a FISH mostra sinais na região pericentroméricas dos pares 3 e 7, umas das regiões descritas por Zattera *et al.*, 2019 como heterocromáticas, demonstrada através de bandamento de C. Em *Physalaemus ephippifer* foi verificada a presença desse elemento clusterizado na região heterocromática do braço curto do cromossomo 3, evidenciando o acúmulo dessas sequências em regiões de heterocromatina (NASCIMENTO; BALDO; LOURENÇO, 2015). O fato de TEs serem encontrados frequentemente em região de heterocromatina (CARDUCCI *et al.*, 2018; TRALDI *et al.*, 2019; VALENTE *et al.*, 2011) pode estar relacionado com a baixa taxa de recombinação nessas regiões, reduzindo assim recombinações ectópicas que pode gerar efeitos deletérios. Outra hipótese é a de que ocorre

maiores eliminação de TEs em regiões onde há uma alta presença de genes, pois esses elementos podem causar mutações deletérias no genoma do hospedeiro, caso haja inserção desse elemento em um gene (HUA-VAN *et al.*, 2005; VALENTE *et al.*, 2011), e assim, seriam removidos por seleção purificadora.

A distribuição dos sinais de hibridação com a sonda Rex1 coincidiram também com marcações observadas em experimentos de FISH com sondas de motivos de microsatélites nos cromossomos de *P. carvalhoi* (Zattera *et al.*, submetido) nas regiões centroméricas dos pares 3 e 7 com o motivo (CA)₁₅, (GA)₁₅, o motivo (GATA)₈ na região terminal do cromossomo 10, (CGC)₁₀ e (CAG)₁₀ na região subterminal dos cromossomos 6 e 10 (ZATTERA *et al.*, submetido). Associação entre TEs e sequências microsatélites não é algo inesperado e uma possível explicação para essa observação podem ser pelo fato que os elementos LINES, em especial o elemento Rex1, possuem repetições microsatélites na extremidade 3' (KAPITONOV; TEMPEL; JURKA, 2009; VOLFF; KORTING; SCHARTL, 2000).

Alguns autores sugerem que microsatélites podem ser sítios preferenciais de inserção de TE (FARRÉ *et al.*, 2011), sendo frequentemente reportadas associações de repetições *in tandem*, principalmente os microsatélites e os elementos de transposição (FARRÉ *et al.*, 2011; GLUGOSKI *et al.*, 2018; KEJNOVSKY *et al.*, 2013; PRIZON *et al.*, 2018). Um exemplo disso pode ser visto em mamíferos, onde foi encontrada uma associação dos elementos *Alu* (classe I, ordem SINE) e *L1* (classe I, ordem LINE) com repetições microsatélites em uma análise genômica. Também, em mapeamento cromossômico por *FISH* em peixes, o elemento Rex3 aparece associado com repetições de microsatélites de (GA)₁₅ e (CA)₁₅ (PRIZON *et al.*, 2018). Uma hipótese para isso seria que essas repetições *in tandem* se assemelhariam muito aos locais de clivagem canônica para esses elementos (FARRÉ *et al.*, 2011). No entanto, não podemos descartar a hipótese de que a presença do microsatélite AAAT em certas regiões pode ter sido derivada de elementos transponíveis degenerados já inseridos no genoma, pois elementos *L1* e *Alu* possuem uma cauda rica em poli(A) e uma região repetida interna *in tandem* (FARRÉ *et al.*, 2011). Essa observação nos leva a uma questão se a associação entre elementos de transposição e repetições microsatélites é uma causa ou um efeito da própria inserção desses elementos em uma determinada região genômica (BUSCHIAZZO; GEMMELL, 2006).

A ausência de sinais de hibridação da sonda Rex1 nos cromossomos de *X. tropicalis* nos levantou questionamentos sobre sua real presença no genoma dessa espécie. Para nos certificarmos que esse elemento está presente realmente no genoma de *X. tropicalis*, utilizamos a região de nucleotídeos correspondente ao fragmento amplificado para a produção da sonda (*contig pipa626994 -757 pb*) como *query* em buscas genômicas por Blastn em *X. tropicalis*. Nessas análises encontramos 12 *hits*, com extensão de alinhamento entre 155-247 pb, apresentando identidade de 82% em média (Tabela 2). Podemos observar que as cópias são poucas e com curta extensão de alinhamento, quando comparamos com os *hits* recuperados do Blastn no genoma de *P. carvalhoi* (mais de 100 *hits* significativos -Anexo I).

Tabela 2: Resumo do resultado das buscas por Blastn de cópias do elemento Rex1 no genoma de *X. tropicalis*. São apresentados os *hits* com alinhamento maior de 100 pb.

ID	Cromossomo	Posição inicial	Posição final	Identidade (%)	Fragmento alinhado (pb)
CM004446.1	cromossomo 4	66374134	66374380	81,78	247
CM004448.1	cromossomo 6	3370868	3371111	81,56	244
CM004448.1	cromossomo 6	80575083	80575216	79,85	134
CM004445.1	cromossomo 3	21362666	21362824	84,28	159
CM004444.1	cromossomo 2	58398792	58398970	83,24	179
CM004444.1	cromossomo 2	62193317	62193149	83,43	169
CM004444.1	cromossomo 2	130188004	130187846	82,39	159
CM004444.1	cromossomo 2	143329590	143329748	83,02	159
CM004449.1	cromossomo 7	64474328	64474463	84,56	136
KV462208.1	unplaced genomic	5010	4854	82,17	157
CM004447.1	cromossomo 5	64963924	64963766	81,13	159
CM004447.1	cromossomo 5	70349253	70349407	81,29	155
CM004447.1	cromossomo 5	71507415	71507282	80,6	134
CM004443.1	cromossomo 1	94587339	94587123	79,26	217

A presença de *hits* significativos no genoma de *X. tropicalis* e a ausência de sinal na hibridização nos indica que realmente existem sequências de Rex1 similares espalhadas no genoma, mas que não conseguimos visualizar no mapeamento físico em experimentos de FISH devido ao próprio poder de resolução do método em questão, à qual requer segmentos de DNA de 10Kb para serem visualizados no microscópio (JIANG *et al.*, 1995) ou que as cópias, mesmo sendo menores, estejam arranjadas em uma mesma porção cromossômica (clusterizadas).

Ainda, o uso de sonda heteróloga deve ter prejudicado o mapeamento desse elemento. De fato, usando como *query* a sequência consenso de Rex1-1_Xt do Repbase contra o genoma de *X. tropicalis*, obtemos um número maior de *hits* significativos (Anexo II). Assim, o grande tempo de divergência entre as duas espécies (cerca de 136 milhões de anos) (HEDGES *et al.*, 2015) resultou na evolução independente desses elementos em cada genoma, de forma que o nível de divergência entre as cópias não permite a eficiência da sonda obtida de *P. carvalhoi* no mapeamento das cópias de Rex1 no genoma de *X. tropicalis*.

Apesar de termos isolados a sequência molecular do elemento RTE e desenhado as sequências dos *primers* diretamente do genoma de *P. carvalhoi* (e amplificados por PCR), não foi possível detectar sinais de hibridação desses elementos no cariótipo dessa espécie (Figura 4). As imagens indicam apenas sinais de hibridação inespecíficas e novos experimentos serão realizados para confirmar tal resultado. De fato, quando usamos a região da sonda de RTE do *contig* pipa84507 como *query* em Blastn contra o genoma de *P. carvalhoi*, um número bem menor de cópias de RTE parece estar presente quando comparadas ao número de cópias de Rex1 (Anexo III). Assim, o baixo número de cópias associado com uma distribuição dispersa no genoma pode ser a explicação para a falta desses sinais de hibridação. Não esperamos encontrar sinais de hibridação de RTE nos cromossomos de *X. tropicalis*, uma vez nossas buscas genômicas (GAMA *et al.*, - em preparação - capítulo 1) reforçam a sugestão de ausência desse elemento no genoma de *X. tropicalis* (SUN *et al.*, 2015).

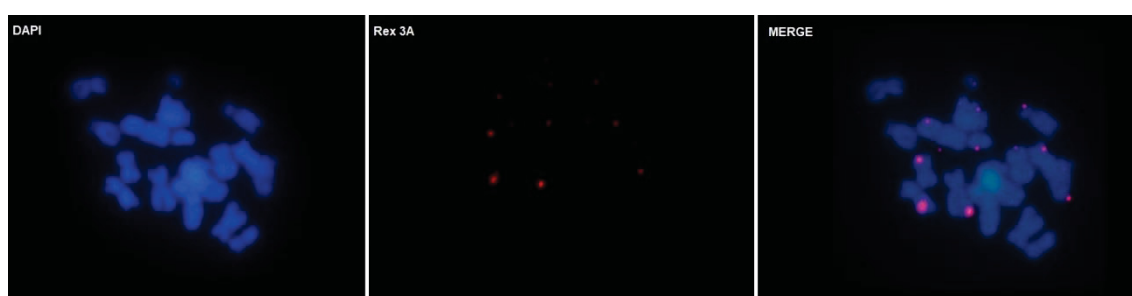


Figura 5. Cromossomos de *P. carvalhoi*, as imagens consecutivamente representam: coloração em DAPI, marcação inespecífica da sonda de RTE marcada com 5-TAMRA-dUTPe merge das duas imagens.

O mapeamento físico das sequências de Rex1 em *P. carvalhoi*, nos mostrou sinais de hibridização, no entanto os resultados obtidos não foram totalmente satisfatórios e novos experimentos deverão ser realizados para que possamos tentar entender e inferir como tais elementos ajudaram a construir a história cromossômica dessa espécie. Os mapeamentos de Rex1 em *X. tropicalis* e RTE em *P. carvalhoi*, também deverão ser refeitos para

confirmarmos se realmente esses elementos estão em baixa densidade e dispersos nesses genomas.

Nossos dados demonstram pela primeira vez a presença e organização cromossômica de cópias de elementos da ordem LINE no cariótipo de Pipídeos e poderão auxiliar no entendimento do papel desses elementos repetitivos na diversificação do cariótipo dessas espécies.

Anexo I- Resumo do resultado de Blastn de Rex1 em *P. carvalhoi*, mostrando os 100 hits mais significativos.

ID contig	Posição inicial	Posição final	Identidade (%)	Fragmento alinhado (pb)
pipa626994	3349	4105	100	757
pipa2249840	938	176	91,53	767
pipa232970	3141	2377	90,2	765
pipa128038	1427	2188	89,5	762
pipa228629	6268	5503	88,9	766
pipa90867	883	122	88,32	762
pipa146549	7248	6482	88,17	769
pipa840159	349	972	90,78	629
pipa14760	2854	3618	87,6	766
pipa621610	990	248	87,29	763
pipa589684	6976	6217	86,87	769
pipa279164	2816	2195	89,76	625
pipa1955760	960	256	87,94	705
pipa38729	8325	7584	86,42	766
pipa282613	2317	1565	86,42	766
pipa934622	944	238	87,9	711
pipa1531500	1	490	93,27	490
pipa1904528	611	1376	86,42	766
pipa61089	2540	3244	87,39	706
pipa816075	10	640	88,59	631
pipa960539	287	996	87,22	712
pipa1044384	1312	606	86,99	707
pipa361966	499	1	91,98	499
pipa157446	621	1	87,83	641
pipa162327	872	128	86,17	745
pipa97430	1	621	88,41	621
pipa376234	215	826	88,44	614
pipa199737	4041	3262	85,17	782
pipa612668	68	680	88,09	613
pipa502411	839	95	85,62	758
pipa2082555	1215	469	85,66	760
pipa1154651	665	149	90,33	517
pipa681028	739	1	85,45	756
pipa2017056	4	617	87,62	614
pipa1115245	610	1	87,87	610
pipa1662693	629	1290	86,87	670
pipa1848362	214	903	86,58	693
pipa1431815	2283	2894	87,75	612
pipa3069408	1	565	88,67	565
pipa249184	609	1	87,68	609

pipa130056	1206	455	84,75	767
pipa1601637	180	811	86,93	635
pipa7338	511	1190	86,03	680
pipa2449107	571	1	88,27	571
pipa382246	2599	3116	89,64	521
pipa362279	1618	2076	91,5	459
pipa2458234	611	43	87,2	586
pipa856627	5499	4791	85,65	711
pipa333137	3460	4038	88,08	579
pipa3037225	25	581	88,53	558
pipa1356483	7	518	89,26	512
pipa1689151	1232	1854	87	623
pipa241134	570	1	87,89	570
pipa916719	1098	550	88,16	549
pipa287655	1016	334	85,82	684
pipa1391078	1116	1641	88,8	527
pipa83718	6883	6127	84,78	762
pipa2708069	470	1	90,27	473
pipa1684431	1	459	90,85	459
pipa1394828	2123	2706	87,39	587
pipa3579133	762	154	86,27	619
pipa522117	597	111	89,53	487
pipa1622112	50	588	88,16	549
pipa205559	1110	1509	92,5	400
pipa241172	2670	3378	84,5	716
pipa1440200	425	911	89,59	490
pipa894182	126	710	86,84	585
pipa89410	1089	465	85,92	625
pipa1032647	1	528	88,07	528
pipa223345	1	496	88,71	496
pipa540588	1	459	90,2	459
pipa40535	495	1	88,91	496
pipa3347553	241	681	90,25	441
pipa392041	4	540	87,52	537
pipa137701	5800	5328	87,7	496
pipa1086520	541	19	87,57	523
pipa1562660	441	1	89,8	441
pipa25747	3357	2854	87,5	504
pipa1939599	25	510	88,07	486
pipa1491642	593	1	85,15	613
pipa2807667	111	585	88,42	475
pipa397554	511	22	87,55	490
pipa164863	436	1015	85,17	580
pipa1011540	502	1	87,48	503

pipa3050503	494	1	87,45	494
pipa211049	780	272	86,64	509
pipa1026150	599	121	87,27	479
pipa121017	38	548	86,32	519
pipa2869788	572	1	85,03	581
pipa1787689	1	403	89,85	404
pipa967835	12	514	86,41	515
pipa540049	1302	838	86,95	475
pipa1176026	9	517	86,33	512
pipa77135	614	50	84,3	586
pipa3679356	501	21	86,94	490
pipa2352962	92	580	87,12	489
pipa1907344	470	1	87,39	476
pipa32614	425	886	86,8	462
pipa935565	471	19	88,13	455

Anexo II- Resumo do resultado de Blastn da cópia canônica Rex1-1_Xt em *X. tropicalis*, mostrando os hits com alinhamento maior de 100 pb.

ID	Cromossomo	Posição inicial	Posição final	Identidade (%)	Fragmento
					alinhado (pb)
CM004443.1	cromossomo 1	82634358	82634230	97,69	130
CM004443.1	cromossomo 1	96913197	96912955	90,8	250
CM004443.1	cromossomo 1	101534564	101534842	95,76	283
CM004443.1	cromossomo 1	124877962	124877863	85,32	109
CM004443.1	cromossomo 1	124878813	124878479	85,56	360
CM004443.1	cromossomo 1	124879663	124878813	92,4	868
CM004443.1	cromossomo 1	140124196	140124348	100	153
CM004443.1	cromossomo 1	140124691	140125492	96,76	803
CM004443.1	cromossomo 1	140126782	140126883	91,59	107
CM004443.1	cromossomo 1	147382063	147382468	84,19	430
CM004443.1	cromossomo 1	154573233	154574774	97,74	1547
CM004444.1	cromossomo 2	43895482	43895364	96,67	120
CM004444.1	cromossomo 2	48359217	48358709	93,04	517
CM004444.1	cromossomo 2	78847585	78847331	85,66	265
CM004444.1	cromossomo 2	97597961	97598280	94,85	330
CM004444.1	cromossomo 2	111930062	111929934	98,45	129
CM004445.1	cromossomo 3	21362747	21362887	80,85	141
CM004445.1	cromossomo 3	53852125	53852252	96,15	130
CM004445.1	cromossomo 3	61259564	61259837	94,53	274
CM004445.1	cromossomo 3	76322692	76322984	86,58	313
CM004445.1	cromossomo 3	118100322	118099970	97,78	360
CM004445.1	cromossomo 3	118100659	118100321	94,89	352
CM004445.1	cromossomo 3	118111207	118111006	99,5	202
CM004445.1	cromossomo 3	124500132	124501156	97,48	1033
CM004446.1	cromossomo 4	12656777	12656496	93,47	291
CM004446.1	cromossomo 4	20169167	20168780	90,72	388
CM004446.1	cromossomo 4	30383184	30383561	95,28	381
CM004446.1	cromossomo 4	30384088	30384642	94,95	555
CM004446.1	cromossomo 4	30384646	30384760	98,26	115
CM004446.1	cromossomo 4	30384758	30385419	93,1	667
CM004446.1	cromossomo 4	51515200	51513709	97,87	1504
CM004446.1	cromossomo 4	51515721	51515201	98,08	521
CM004446.1	cromossomo 4	65355203	65355321	91,47	129
CM004446.1	cromossomo 4	82121314	82121099	93,52	216
CM004446.1	cromossomo 4	82123719	82123524	97,46	197
CM004446.1	cromossomo 4	82125519	82124893	94,44	630
CM004446.1	cromossomo 4	87025809	87026028	95,5	222
CM004447.1	cromossomo 5	35306693	35306931	94,14	239

CM004447.1	cromossomo 5	37378229	37379080	94,87	858
CM004447.1	cromossomo 5	70348898	70349023	80,62	129
CM004447.1	cromossomo 5	98927715	98927594	92,62	122
CM004448.1	cromossomo 6	15954463	15954335	96,12	129
CM004448.1	cromossomo 6	39696794	39697320	95,68	532
CM004448.1	cromossomo 6	39697314	39698086	95,71	793
CM004448.1	cromossomo 6	39698094	39698400	96,74	307
CM004448.1	cromossomo 6	39698403	39698533	90,08	131
CM004448.1	cromossomo 6	41998167	41998584	96,89	418
CM004448.1	cromossomo 6	41998588	42000471	96,64	1902
CM004448.1	cromossomo 6	49975282	49975862	96,43	589
CM004448.1	cromossomo 6	98284007	98283503	93,8	516
CM004448.1	cromossomo 6	129873157	129872733	89,11	459
CM004449.1	cromossomo 7	21186919	21187130	98,12	213
CM004449.1	cromossomo 7	71866160	71866786	88,26	639
CM004449.1	cromossomo 7	71866788	71866946	93,08	159
CM004449.1	cromossomo 7	71868148	71868494	89,58	355
CM004449.1	cromossomo 7	71868629	71868853	92,44	225
CM004449.1	cromossomo 7	71868862	71869560	83,61	726
CM004449.1	cromossomo 7	74723658	74723485	96,07	178
CM004449.1	cromossomo 7	74723889	74723655	96,62	237
CM004449.1	cromossomo 7	74725579	74723968	96,2	1632
CM004449.1	cromossomo 7	85285621	85285302	86,73	324
CM004450.1	cromossomo 8	42419102	42418291	95,96	816
CM004450.1	cromossomo 8	42419352	42419114	95,4	239
CM004450.1	cromossomo 8	42419525	42419352	95,43	175
CM004450.1	cromossomo 8	42420663	42420399	96,98	265
CM004450.1	cromossomo 8	49654028	49654299	96,72	274
CM004450.1	cromossomo 8	63825878	63825778	85,15	101
CM004450.1	cromossomo 8	103182884	103183347	90,87	482
CM004451.1	cromossomo 9	10470388	10470653	89,57	278
CM004451.1	cromossomo 9	33550517	33550777	98,09	262
CM004451.1	cromossomo 9	35100283	35099939	83,47	363
CM004452.1	cromossomo 10	5161144	5161016	96,15	130
KV460357.1	unplaced	690333	690222	100	112
KV460417.1	unplaced	286914	286058	88,8	884
KV460417.1	unplaced	399431	400023	96,3	595
KV460417.1	unplaced	400025	401724	96,16	1717
KV460432.1	unplaced	8469	7560	96,49	912
KV460494.1	unplaced	246814	246980	98,2	167
KV460494.1	unplaced	247621	247841	94,12	221
KV460494.1	unplaced	248308	249170	93,62	878

Anexo III- Resumo do resultado de Blastn de RTE em *P. carvalhoi*, mostrando todos os hits significativos com alinhamento acima de 100 pb.

ID <i>contig</i>	Posição inicial	Posição final	Identidade (%)	Fragmento alinhado
pipa1532542	1751	1227	94.86	525
pipa391493	3425	3948	94.27	524
pipa470793	3710	4232	94.46	523
pipa542043	343	862	94.23	520
pipa511654	894	371	93.70	524
pipa1852226	933	409	93.71	525
pipa1042351	28	551	93.32	524
pipa449877	1316	793	93.32	524
pipa1610849	14	538	93.33	525
pipa434492	1864	1343	93.32	524
pipa148807	6793	7315	92.93	523
pipa352876	524	105	93.33	420
pipa3074236	409	1	92.67	409
pipa4660808	288	607	94.38	320
pipa1363387	453	175	95.34	279
pipa501376	329	73	93.46	260
pipa2483800	832	567	92.48	266
pipa4068245	246	1	93.09	246
pipa501376	543	331	94.37	213
pipa4278853	311	109	93.66	205
pipa601090	2704	2890	94.65	187
pipa186797	3321	3525	91.47	211
pipa1001397	830	974	96.55	145
pipa152451	207	323	96.58	117

3. CONCLUSÕES E CONSIDERAÇÕES FINAIS

- Foi possível, por meio do *pipeline* PIRATE, obter pela primeira vez o panorama global do conteúdo de TEs e sequências repetitivas presentes no genoma de *P. carvalhoi*, mesmo que o conteúdo seja subestimado devido a filtragem dos *contigs*. Sabe-se que estão presentes pelo menos 22 superfamílias de TEs nesse genoma.
- Foram recuperadas as sequências dos elementos genericamente denominados Rex (Rex1, Rex3, Rex5 e Rex6) no genoma dos anuros disponível no GenBank e as relações filogenéticas entre essas sequências foram inferidas. Com isso, foi possível estabelecer que os elementos denominados Rex não foram recuperados como um grupo monofilético
- Refutamos suposta transferência horizontal do elemento RTE entre Anuros e Hirudinea. Apesar da inconsistência verificada pelo agrupamento dessas sequências na árvore, o teste indicou que possivelmente essas sequências foram herdadas por transmissão vertical. Um estudo especificamente delineado para acessar essa questão poderá auxiliar a entender melhor a origem da incomum similaridade de sequências de RTE encontrada entre as espécies
- O mapeamento cromossômico do elemento Rex1, realizado primeira vez em anuros da família Pipidae, revelou clusters de sequências preferencialmente em porções heterocromáticas dos cromossomos, com associação com outros elementos repetitivos, como motivos microssatélite previamente mapeados nesses cromossomos. Essa organização cromossômica se assemelha à outros relatos desse elemento em peixes e em uma espécie de anuro.
- Existem poucos estudos a respeito de genômica de anuros, principalmente investigando a classificação de tais elementos, assim, as contribuições apresentadas nesse trabalho abrangem também informações a respeito da classificação desses componentes genômicos. Nossa análise filogenética demonstrou os elementos da ordem LINEs nos seis grandes grupos apresentados por Kapitonov *et al* (2009) possivelmente não reflete as relações filogenéticas entre os clados.
- Elementos de transposição são importantes efetores na evolução dos genomas. Assim, esse trabalho pode ser considerado um dos primeiros passos para o entendimento da evolução e organização desses elementos em anuros.

REFERÊNCIAS

- ALTSCHUL, S. F. *et al.* Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, n. 17, p. 3389–3402, 1997.
- BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, n. 1, p. 4–9, 2015.
- BARTOLOMÉ, C.; BELLO, X.; MASIDE, X. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. **Genome Biology**, v. 10, n. 2, p. 1–11, 2009.
- BENNETZEN, J. L. Transposable elements , gene creation and genome rearrangement in flowering plants. **Current Opinion in Genetics & Development**, v. 15, p. 621–627, 2005.
- BEREZIKOV, E.; BUCHETON, A.; BUSSEAU, I. A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of *Drosophila melanogaster*. **Genome biology**, v. 1, n. 6, p. 1–15, 2000.
- BIÉMONT, C. A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution ´. **Genetics Society of America**, v. 186, p. 1085–1093, 2010.
- BIÉMONT, C.; VIEIRA, C. Junk DNA as an evolutionary force. **Nature**, v. 443, n. October, p. 521–524, 2006.
- BIESSMANN, H. *et al.* Addition of Telomere-Associated HeT DNA Sequences ‘ Heals ’ Broken Chromosome Ends in *Drosophila*. **cell**, v. 61, p. 663–673, 1990.
- BOUNEAU, L. *et al.* An active non-LTR retrotransposon with tandem structure in the compact genome of pufferfish *Tetraodon nigroviridis*. **Genome Research**, v. 13, n. 7, p. 1686–1695, 2003.
- BRUSCHI, D. P. *et al.* Cytogenetic analysis of two species in the *Phyllomedusa hypochondrialis* group (Anura , Hylidae). **Hereditas**, v. 40, p. 34–40, 2012.
- BURKE, W. D. *et al.* Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. **Molecular Biology and Evolution**, v. 19, n. 5, p. 619–630, 2002.
- BUSCHIAZZO, E.; GEMMELL, N. J. The rise , fall and renaissance of microsatellites in eukaryotic genomes. **BioEssays**, v. 28, p. 1040–1050, 2006.
- CANNATELLA, D. C.; TRUEB, L. Evolution of pipoid frogs : intergeneric relationships of the aquatic frog family Pipidae (Anura). **Zoological Journal of the Linnean Society**, v. 94, p. 1–38, 1988.
- CAPELLA-GUTIÉRREZ, S.; SILLA-MARTÍNEZ, J. M.; GABALDÓN, T. trimAl : a tool for automated alignment trimming in large-scale phylogenetic analyses. **Bioinformatics Applications note**, v. 25, n. 15, p. 1972–1973, 2009.
- CAPY, P.; ANXOLABÉHÈRE, D.; LANGIN, T. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? **Trends in Genetics**, v. 10, n. 1, p. 8–11, 1994.
- CARDUCCI, F. *et al.* Rex Retroelements and Teleost Teleost Genomes : An Overview.

International Journal of Molecular Sciences, v. 19, p. 1–15, 2018.

CHALOPIN, D. *et al.* Evolutionary Active Transposable Elements in the Genome of the Coelacanth. **Journal of Experimental Zoology**, v. 322B, p. 322–333, 2013.

CHALOPIN, D. *et al.* Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. **Genome Biology and Evolution**, v. 7, n. 2, p. 567–580, 2015.

DASILVA, C. *et al.* Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 21, p. 13636–13641, 2002.

DELPRAT, A. *et al.* The Transposon Galileo Generates Natural Chromosomal Inversions in *Drosophila* by Ectopic Recombination. **PLoS ONE**, v. 4, n. 11, p. 1–13, 2009.

EDWARDS, R. J. *et al.* Draft genome assembly of the invasive cane toad, *Rhinella marina*. **GigaScience**, v. 7, p. 1–13, 2018.

EICKBUSH, T. H.; MALIK, H. S. Origins and Evolution of Retrotransposons. In: CRAIG, N. L. *et al.* (Eds.). **Mobile DNA II**. Washington: American Society of Microbiology Press, 2002. p. 1111–1144.

FARRÉ, M. *et al.* Assessing the Role of Tandem Repeats in Shaping the Genomic Architecture of Great Apes. **PLoS ONE**, v. 6, n. 11, p. 1–12, 2011.

FAVARATO, R. M. *et al.* Chromosomal mapping of transposable elements of the rex family in the bristlenose catfish, *Ancistrus (siluriformes, loricariidae)*, from the amazonian region. **Journal of Heredity**, v. 108, n. 3, p. 254–261, 2017.

FESCHOTTE, C.; PRITHAM, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. **The Annual Review of Genetics**, v. 41, p. 331–368, 2007.

FINNEGAN, D. J. Eukaryotic Transposable elements and Genome Evolution. **Trends in genetics**, v. 5, n. 4, p. 103–107, 1989.

FISCHER, C. *et al.* Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish *Tetraodon nigroviridis*. **Gene**, v. 336, n. 2, p. 175–183, 2004.

GARRETT, J. E.; CARROLL, D. Txl: a Transposable Element from *Xenopus laevis* with Some Unusual Properties. **Molecular and Cellular Biology**, v. 6, n. 3, p. 933–941, 1986.

GARRETT, J. E.; KNUTZON, D. S.; CARROLL, D. Composite Transposable Elements in the *Xenopus laevis* Genome. **Molecular and Cellular Biology**, v. 9, n. 7, p. 3018–3027, 1989.

GLUGOSKI, L. *et al.* Co-located hAT transposable element and 5S rDNA in an interstitial telomeric sequence suggest the formation of Robertsonian fusion in armored cat fish. **Gene**, v. 650, n. February, p. 49–54, 2018.

GRAINGER, R. *Xenopus tropicalis* as a Model Organismo for Genetics and Genomics: Past, Present and Future. In: **Xenopus protocol: methods in Molecular Biology**. [s.l.: s.n.]. v.

917p. 279–292.

GRAY, Y. H. M. It takes two transposons to tango. **Trends in Genetics**, v. 16, n. 10, p. 461–468, 2000.

HAN, J. S. Non-long terminal repeat (non-LTR) retrotransposons: Mechanisms, recent developments, and unanswered questions. **Mobile DNA**, v. 1, n. 1, p. 1–12, 2010.

HEDGES, S. B. *et al.* Tree of life reveals clock-like speciation and diversification. **Molecular Biology and Evolution**, v. 32, n. 4, p. 835–845, 2015.

HELLSTEN, U. *et al.* The Genome of the western Clawed Frog *Xenopus tropicalis*. **Science**, v. 328, n. 5978, p. 633–636, 2010.

HIKOSAKA, A.; YOKOUCHI, E.; KAWAHARA, A. Extensive Amplification and Transposition of a Novel Repetitive Element , Xstir , Together with Its Terminal Inverted Repeat in the Evolution of *Xenopus*. **Journal of Molecular Evolution**, v. 51, p. 554–564, 2000.

HUA-VAN, A. *et al.* Abundance, distribution and dynamics of retrotransposable elements and transposons: Similarities and differences. **Cytogenetic and Genome Research**, v. 110, n. 1–4, p. 426–440, 2005.

IRISARRI, I. *et al.* Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs , family Pipidae. **BMC Evolutionary Biology**, v. 11, n. 114, p. 1–10, 2011.

JIANG, J. *et al.* Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 10, p. 4487–4491, 1995.

JIANG, N. *et al.* Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). **Current Opinion in Plant Biology**, v. 7, p. 115–119, 2004.

JURKA, J. *et al.* Repbase Update , a database of eukaryotic repetitive elements. **Cytogenetic and Genome Research**, v. 110, p. 462–467, 2005.

JURKA, J. *et al.* Repetitive Sequences in Complex Genomes : Structure and Evolution. **Annual Review of Genomics and Human Genetics**, v. 8, p. 241–259, 2007.

KAMINKER, J. S. *et al.* The transposable elements of the *Drosophila melanogaster* euchromatin : a genomics perspective. **Genome Biology**, v. 3, p. 1–20, 2002.

KAPITONOV, V. V.; TEMPEL, S.; JURKA, J. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. **Gene**, v. 448, n. 2, p. 207–213, 2009.

KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature genetics**, v. 9, p. 411–412, 2008.

KATOH, K.; ROZEWICKI, J.; YAMADA, K. D. MAFFT online service : multiple sequence alignment , interactive sequence choice and visualization. **Briefings in Bioinformatics**, p. 1–7, 2017.

- KEINATH, M. C. *et al.* Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. **Scientific Reports**, v. 5, n. 16413, p. 1–13, 2015.
- KEJNOVSKY, E. *et al.* Expansion of Microsatellites on Evolutionary Young Y Chromosome. **PLoS ONE**, v. 8, n. 1, p. 1–10, 2013.
- KIBBE, W. A. OligoCalc : an online oligonucleotide properties calculator. **Nucleic Acids Research**, v. 35, p. 43–46, 2007.
- KIM, J. M. *et al.* Transposable Elements and Genome Organization : A Comprehensive Survey of Retrotransposons Revealed by the Complete *Saccharomyces cerevisiae* Genome Sequence. **Genome research**, v. 8, n. 515, p. 464–478, 1998.
- KOJIMA, K. K.; FUJIWARA, H. An extraordinary retrotransposon family encoding dual endonucleases. **Genome Research**, v. 15, n. 8, p. 1106–1117, 2005.
- KUMAR, S. *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 2018.
- LARKIN, D. M. *et al.* Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. **Genome Biology and Evolution**, p. 1–9, 2006.
- LEACH, D. R. F. Long DNA palindromes cruciform structures, genetic instability and secondary structure repairs. **BioEssays**, v. 16, n. 12, p. 893–900, 1994.
- LEPETIT, D. *et al.* Glider and Vision : two new families of miniature inverted-repeat transposable elements in *Xenopus laevis* genome. **Genetica**, v. 108, p. 163–169, 2000.
- LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL) v4 : recent updates and. **Nucleic Acids Research**, v. 47, n. April, p. 256–259, 2019.
- LIEDTKE, H. C. *et al.* Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. **Nature Ecology & Evolution**, v. 2, p. 1792–1799, 2018.
- LIM, J. K.; SIMMONS, M. J. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. **BioEssays**, v. 16, n. 4, p. 269–274, 1994.
- LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. . The Repetitive DNA Content of Eukaryotic Genomes. In: GARRIDO-RAMOS, M. . (Ed.). . **Repetitive DNA**. 7. ed. [s.l.] Genome Dynamic, karger, 2012. v. 7p. 1–28.
- LOVŠIN, N.; GUBENŠEK, F.; KORDIŠ, D. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in deuterostomia. **Molecular Biology and Evolution**, v. 18, n. 12, p. 2213–2224, 2001.
- LUDWIG, A.; VALENTE, V. L. D. S.; LORETO, E. L. S. Multiple invasions of Errantivirus in the genus *Drosophila*. **Insect Molecular Biology**, v. 17, n. 2, p. 113–124, 2008.
- MADEIRA, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. **Nucleic acids research**, v. 47, n. W1, p. W636—W641, 2019.
- MALIK, H. S.; BURKE, W. D.; EICKBUSH, T. H. The Age and Evolution of Non-LTR Retrotransposable Elements. **Molecular Biology and Evolution**, v. 16, n. 6, p. 793–805,

1999.

MALIK, H. S.; EICKBUSH, T. H. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. **Molecular Biology and Evolution**, v. 15, n. 9, p. 1123–1134, 1998.

MARCHLER-BAUER, A.; BRYANT, S. H. CD-Search : protein domain annotations on the fly. **Nucleic Acids Research**, v. 32, n. W327–W331, p. 327–331, 2004.

MILLER, M. A.; PFEIFFER, W.; SCHWARTZ, T. **Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees**. Gateway Computing Environments Workshop (GCE), New Orleans, LA. **Anais...**2010

NASCIMENTO, J.; BALDO, D.; LOURENÇO, L. B. First insights on the retroelement Rex1 in the cytogenetics of frogs. **Molecular Cytogenetics**, v. 8, n. 86, p. 1–9, 2015.

NOVIKOVA, O. *et al.* CR1 clade of non-LTR retrotransposons from Maculinea butterflies (Lepidoptera: Lycaenidae): Evidence for recent horizontal transmission. **BMC Evolutionary Biology**, v. 7, p. 1–19, 2007.

NOVIKOVA, O.; FET, V.; BLINOV, A. Non-LTR retrotransposons in fungi. **Functional and Integrative Genomics**, v. 9, n. 1, p. 27–42, 2009.

OZOUF-COSTAZ, C. *et al.* Genome dynamics and chromosomal localization of the non-LTR retrotransposons Rex1 and Rex3 in Antarctic fish. **Antarctic Science**, v. 16, n. 1, p. 51–57, 2004.

PARDUE, M. *et al.* Two retrotransposons maintain telomeres in *Drosophila*. **Chromosome Research**, v. 13, n. 5, p. 443–453, 2005.

PARKER, H. G. *et al.* An Expressed Fgf3 Retrotransposon Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. **Science**, v. 325, p. 995–997, 2009.

PRIZON, A. C. *et al.* Chromosome Spreading of the Retrotransposable Rex -3 Element and Microsatellite Repeats in Karyotypes. **ZEBRAFISH**, v. 00, n. 00, p. 1–11, 2018.

PUTNAM, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. **Science**, v. 317, n. 5834, p. 86–94, 2007.

SEBERG, O.; PETERSEN, G. A unified classification system for eukaryotic transposable elements should reflect their phylogeny. **Nature**, v. 224, p. 1, 2009.

SESSION, A. M. *et al.* Genome evolution in the allotetraploid frog *Xenopus laevis*. **Nature**, v. 538, n. 7625, p. 336–343, 2016.

SHAO, F.; HAN, M.; PENG, Z. Evolution and diversity of transposable elements in fish genomes. **Scientific Reports**, v. 9, p. 1–8, 2019.

SILVA, E. *et al.* Distribution extension and geographic distribution map in the southeast of Brazil. **Journal of species lists and distribution**, v. 6, n. 3, p. 451–453, 2010.

SILVA, J. C.; LORETO, E. L.; CLARK, J. B. Factors that affect the horizontal transfer of transposable elements. **Current Issues in Molecular Biology**, v. 6, n. 1, p. 57–72, 2004.

- SILVA, J. C.; MARGARET, G. K. Horizontal transfer and selection in the evolution of P elements. **Molecular Biology and Evolution**, v. 17, n. 10, p. 1542–1557, 2000.
- SOTERO-CAIO, C. G. *et al.* Evolution and Diversity of Transposable Elements in Vertebrate Genomes. **Genome Biology**, v. 9, n. 1, p. 161–177, 2017.
- STAMATAKIS, A. RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics Applications Note**, v. 30, n. 9, p. 1312–1313, 2014.
- SUN, Y. *et al.* Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. **Proceedings of the National Academy of Sciences of The United States of America**, n. 1257, p. 1262, 2015.
- TINSLEY, R. C. The evidence from parasite relationships for the evolutionary status of xenopus (*Anura pipidae*). **Monitore Zoologico Italiano, Supplemento**, v. 15, n. 1, p. 367–385, 1981.
- TRALDI, J. B. *et al.* Chromosomal distribution of the retroelements Rex1, Rex3 and Rex6 in species of the genus *Harttia* and *Hypostomus* (Siluriformes: Loricariidae). **Neotropical Ichthyology**, v. 17, n. 2, p. 1–10, 2019.
- VALENTE, G. T. *et al.* Cytogenetic mapping of the retroelements Rex1, Rex3 and Rex6 among cichlid Fish: New insights on the chromosomal distribution of transposable elements. **Cytogenetic and Genome Research**, v. 133, n. 1, p. 34–42, 2011.
- VOLFF, J. *et al.* The Non-LTR Retrotransposon Rex3 from the Fish *Xiphophorus* is Widespread Among Teleosts. **Molecular Biology and Evolution**, v. 16, n. 11, p. 1427–1438, 1999.
- VOLFF, J. *et al.* Non-LTR Retrotransposons Encoding a Restriction Enzyme-Like Endonuclease in Vertebrates. **Molecular Biology and Evolution**, v. 52, p. 351–360, 2001.
- VOLFF, J.; KORTING, C.; SCHARTL, M. Multiple Lineages of the Non-LTR Retrotransposon Rex1 with Varying Success in Invading Fish Genomes. **Molecular Biology and Evolution**, v. 17, n. 11, p. 1673–1684, 2000.
- WALKERL, E. L. *et al.* Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. **The EMBO Journal**, v. 14, n. 10, p. 2350–2363, 1995.
- WALLAU, G. L.; VIEIRA, C.; LORETO, É. L. S. Genetic exchange in eukaryotes through horizontal transfer: Connected by the mobilome. **Mobile DNA**, v. 9, n. 1, p. 1–16, 2018.
- WICKER, T. *et al.* A unified classification system for eukaryotic transposable elements. **Nature genetics**, v. 8, p. 973–982, 2007.
- WILDER, J.; HOLLOCHER, H. Mobile Elements and the Genesis of Microsatellites in Dipterans. **Molecular Biology and Evolution**, v. 18, n. 3, p. 384–392, 2001.
- XIA, X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1550–1552, 2018.
- ZATTTERA, M. L. **Organização cromossômica de DNAs repetitivos no cariótipo de *Pipa carvalhoi* (Miranda-Ribeiro, 1937) (Anura: Pipidae)**. 2019. 86 f. Dissertação (Mestrado em

Genética) – Setor de Biológicas, Universidade Federal do Paraná, Curitiba, 2019.