

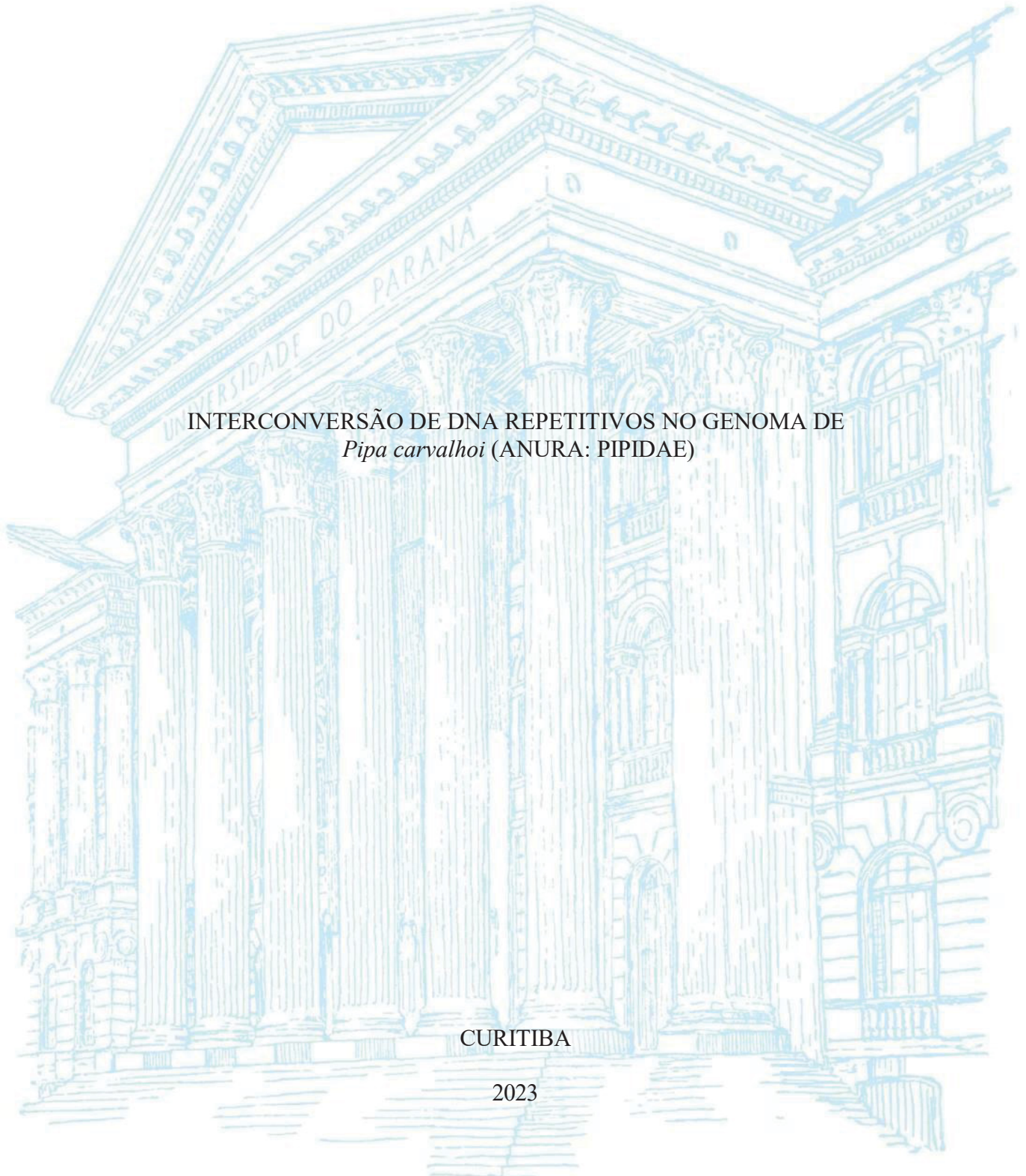
UNIVERSIDADE FEDERAL DO PARANÁ

MICHELLE LOUISE ZATTERA

INTERCONVERSÃO DE DNA REPETITIVOS NO GENOMA DE  
*Pipa carvalhoi* (ANURA: PIPIDAE)

CURITIBA

2023



MICHELLE LOUISE ZATTERA

INTERCONVERSÃO DE DNA REPETITIVOS NO GENOMA DE  
*Pipa carvalhoi* (ANURA: PIPIDAE)

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

Orientador: Prof. Dr. Daniel Pacheco Bruschi

CURITIBA

2023

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

Zattera, Michelle Louise, 1994-.

Interconversão de DNAs repetitivos no genoma de *Pipa carvalhoi*  
(ANURA: PIPIDAE). / Michelle Louise Zattera. – Curitiba, 2023.  
1 recurso on-line : PDF.

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

Orientador: Prof. Dr. Daniel Pacheco Bruschi.

1. DNA. 2. *Xenopus*. 3. Elementos de DNA Transponíveis. 4. Pipidae. I. Bruschi, Daniel Pacheco, 1987-. II. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética. III. Título.

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **MICHELLE LOUISE ZATTERA** intitulada: **INTERCONVERSÃO DE DNAs REPETITIVOS NO GENOMA DE *Pipa carvalhoi* (ANURA: PIPIDAE)**

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

CURITIBA, 18 de Dezembro de 2023.

Assinatura Eletrônica

19/12/2023 12:09:04.0

DANIEL PACHECO BRUSCHI

Presidente da Banca Examinadora

Assinatura Eletrônica

19/12/2023 10:22:20.0

STENIO EDER VITTORAZZI

Avaliador Externo (50002015)

Assinatura Eletrônica

19/12/2023 12:46:43.0

KALEB PRETTO GATTO

Avaliador Externo (UNIVERSIDADE ESTADUAL DE SÃO PAULO)

Assinatura Eletrônica

19/12/2023 10:58:05.0

IRIS HASS

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

*Dedico essa tese à minha avó Mercedes, in memoriam.*

*Sua presença, amor, força e teimosia constantes foram – e seguem sendo – a base da nossa família. Obrigada por sempre me ajudar a tornar meus sonhos realidade.*

## AGRADECIMENTOS

Agradeço, primeiramente, a minha família. A razão de tudo que faço. Amanda, Tia Ari e Lara, obrigada por compreenderem as infinitas horas trancada em meu quarto e pelo apoio constante, fosse em palavras e ou em ações. Foram vocês que me mantiveram em pé quando tudo estava desmoronando. Mãe, meu alicerce para tudo e porto seguro que sempre terei para onde voltar, obrigada por me ensinar que sempre vale a pena seguir lutando por aquilo que desejamos. Sem o amor constante (e os puxões de orelha necessários) eu não chegaria aonde cheguei. Você é minha estrela guia e meu exemplo para tudo. Obrigada por existir.

Aos meus amigos e companheiros de laboratório ao longo desses anos. Johnny, Camilla, Rafael, Joana, Fernanda Gatto, Amanda, César, Vini, Fer, Iris, Enzo, Luíz, Tainá e Ketelen. As conversas (e fofocas) regadas a muito café eram um bálsamo em meio ao caos. Obrigada por todo o apoio, riso e incentivo no meio do caminho.

Meus agradecimentos sinceros e com profunda gratidão ao meu orientador Prof. Dr. Daniel Pacheco Bruschi, o qual devo mais do que um dia serei capaz de retribuir. Ao longo dos últimos sete anos, descobri a importância de uma orientação pautada no respeito à pesquisa, no amor pela ciência e, principalmente, na importância do respeito ao lado humano que caminha ao lado da vida acadêmica. Sua orientação desde o mestrado, e todo o apoio que recebi ao longo destes quase cinco anos de doutorado me mantiveram seguindo todas as vezes que eu pensava em, talvez, desistir. Obrigada por confiar em mim por tantos anos, por me abrir portas que eu jamais pensei que poderiam ser abertas, e por ser sempre tão presente em cada etapa do caminho.

Agradeço também ao MSc. Elverson Soares de Melo, por toda a sua ajuda e paciência com as análises de bioinformática, tão importantes e fundamentais para a conclusão de minha tese.

À UFPR, por ter sido meu lar pelos últimos sete anos. Ter tido o privilégio de realizar meu mestrado e doutorado em uma instituição que tanto amo e respeito, é algo

que me enche de orgulho e alegria. Agradeço também à CAPES, pela concessão da bolsa de auxílio a pesquisa que permitiu desenvolver essa pesquisa ao longo dos anos com dedicação exclusiva, principalmente em um período tão delicado quanto o da pandemia de COVID-19.

Sinto que existem inúmeras pessoas, de perto e de longe, que me permitiram chegar até este momento. Então, deixo meu agradecimento a todos que de alguma forma contribuíram para isso e me ajudaram a crescer como pessoa. A Michelle que hoje escreve essa tese é, também, resultado do somatório de todos vocês. Obrigada, do mais profundo do meu coração.

Por último, e mais importante, à minha Nonna (*in memoriam*). Essa é minha primeira conquista sem a senhora aqui para ver. Ainda que grande parte das outras você também não tenha compreendido muito bem, o abraço sincero e o sorriso no rosto sempre receberam cada uma das minhas vitórias. A senhora ainda é – e sempre será - meu maior referencial de força. Obrigada por continuar me inspirando.

*"Você não pode passar um único dia sem causar um impacto no mundo ao seu redor. Suas ações fazem a diferença, e você precisa que decidir que tipo de diferença deseja fazer."*

***Jane Goodall***



## RESUMO

Os elementos transponíveis (TEs) se apresentam como um dos maiores componentes dos genomas eucariotos, exercendo papéis importantes na regulação da expressão gênica, participando de rearranjos cromossômicos através do seu processo de mobilidade, além de serem possíveis origens para novas sequências repetitivas em tandem. Logo, compreender a dinâmica e a influência destes elementos na geração de sequências repetitivas é crucial para a compreensão do papel desempenhado por eles na evolução genômica. O objetivo da presente tese consistiu em entender o processo de surgimento, manutenção e evolução de sequências de DNAs repetitivos organizados “em tandem” no genoma de *P. carvalhoi* e *X. tropicalis*, com foco nos elementos de transposon de DNA hAT. Em um primeiro momento, foi realizada uma revisão de literatura acerca dos casos já documentados de elementos transponíveis como ponto de origem de novas sequências em tandem, buscando compreender melhor esta relação, através da avaliação da importância do ciclo de vida dos TEs, bem como sua estrutura molecular e a distribuição cromossômica de suas cópias. Posteriormente, utilizando abordagens bioinformáticas, o foco das análises se deu no elemento hAT encontrado aleatoriamente no genoma de *Pipa carvalhoi*, isolado, clonado e mapeado cromossomicamente com o uso de primers heterólogos para Rex-3. A comparação com dados de *Xenopus tropicalis*, espécie irmã do objeto de estudo, revelou importantes percepções sobre a estabilidade genômica e fragmentação desses elementos. Analisando a sequência do clone obtido em *P. carvalhoi*, em análises de similaridade com sequências de *X. tropicalis* foi possível verificar a presença do elemento genômico hAT\_10 em *Pipa carvalhoi*, destacando ainda presença de repetições próximas às regiões de ORFs do elemento como possíveis de regiões de satélites dentro de ambas as espécies. A ausência de ORFs intactas nos elementos de hAT, destacam que estes elementos podem não estar ativos dentro destes genomas. Em última instância, discute-se se o local de inserção dos elementos transponíveis nos genomas pode ser uma forma de mitigar conflitos genéticos com o “hospedeiro”. Estas descobertas destacam a presença de sequências repetitivas inseridas dentro da estrutura dos elementos transponíveis de hAT, um fenômeno documentado pela primeira vez no genoma de *Pipa carvalhoi*. Além disso, este é o primeiro registro deste elemento específico na espécie, proporcionando uma contribuição significativa para a compreensão da dinâmica dos elementos transponíveis dentro do gênero de *Pipa* como um todo.

**Palavras-chave:** Elementos transponíveis, DNA satélites, elementos hAT, *Xenopus tropicalis*, *Pipa carvalhoi*.

## ABSTRACT

Transposable elements (TEs) stand out as one of the largest components of eukaryotic genomes, playing crucial roles in gene expression regulation, participating in chromosomal rearrangements through their mobility process, and serving as potential sources for new tandemly repeated sequences. Therefore, understanding the dynamics and influence of these elements in generating repetitive sequences is crucial for comprehending their role in genomic evolution. The objective of this thesis was to understand the emergence, maintenance, and evolution of tandemly organized repetitive DNA sequences in the genomes of *P. carvalhoi* and *X. tropicalis*, with a focus on DNA transposon elements of the hAT family. Initially, a literature review explored documented cases of transposable elements serving as the origin of new tandem sequences, aiming to better understand this relationship, evaluating the importance of TE life cycles, molecular structure, and chromosomal distribution. Subsequently, using bioinformatic approaches, the analysis focused on the hAT element randomly found in the genome of *Pipa carvalhoi*, isolated, cloned, and chromosomally mapped using heterologous primers for Rex-3. Comparison with *Xenopus tropicalis* data, a sister species, revealed insights into the genomic stability and fragmentation of these elements. Analyzing the sequence of the clone obtained in *P. carvalhoi*, similarity analyses with *X. tropicalis* sequences confirmed the presence of the genomic hAT\_10 element in *Pipa carvalhoi*, highlighting the presence of repeats near the element's ORF regions as potential satellite regions in both species. The absence of intact ORFs in hAT elements suggests that these elements may not be active within these genomes. Ultimately, the discussion revolves around whether the insertion site of transposable elements in genomes might be a way to mitigate genetic conflicts with the host. These findings reveal the presence of repetitive sequences inserted within the structure of hAT transposable elements; a phenomenon documented for the first time in the genome of *Pipa carvalhoi*. Additionally, this marks the first record of this specific element in the species, offering a significant contribution to understanding the dynamics of transposable elements within the *Pipa* genus.

**Keywords:** Transposable elements, satDNA, hAT elements, *Xenopus tropicalis*, *Pipa carvalhoi*.

## LISTA DE FIGURAS

### **CAPÍTULO 1: Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome**

**Figure 1:** Life cycle of a transposable element, using LTR transposons with their respective Open Reading Frames (ORFs) as an example. **T0**—simple copy of the element in the genome; **T1—Invasion:** the chromosomal locus acquires at least one repeat unit; **T2—Amplification:** an increase in the number of copies present in the genome, either by burst events (in the case of class I elements) or the repair or homologous recombination of double-stranded DNA (class II); **T3—Maturation:** the elements are inactivated or silenced through epigenetic silencing, piRNAs, DNA methylation or other mechanisms, such as mutations; **T4—Death or degeneration:** the elements may be either eliminated from the genome, undergo molecular domestication, begin to exercise new functions or give rise to new repetitive sequences, such as tandem repeats. **LTR** = Long Terminal Repeat; **ORF** = Open Reading Frame.

**Figure 2:** Schematic diagram of the different regions of transposable elements that may provide the starting point of origin for new micro-, mini- or satellite DNA. **LTR** (Long Terminal Repeat); **GAG** (GAG domain); **POL** (Reverse Transcriptase); **ENV** (Envelope Protein); **UTR** (Untranslated Region); **ORF** (Open Reading Frame); **TIR** (Terminal Inverted Repeats).

**Figure 3:** Schematic diagram showing how the Non-Homologous End-Joining (NHEJ) and Non-Allelic Homologous Recombination (NAHR) DNA repair mechanisms contribute to the expansion of tandem repeats from a Transposable Element. **TSD** (Target Site Duplication); **TIR** (Terminal Inverted Repeats); **ORF** (Open Reading Frame).

### **CAPÍTULO 2: Elementos de transposição como reservatórios de unidades repetitivas durante o surgimento de um novo DNA satélite: o caso dos elementos hAT no genoma de *P. carvalhoi* e de *X. tropicalis***

**Figura 1:** Cromossomos metafásicos de *P. carvalhoi* submetidos à FISH com sonda isolada a partir dos *primers* de Rex 3.

**Figura 2:** Dendrograma das sequências de hAT inferida pelo critério de ML (Maximum Likelihood) a partir das cópias com maior nível de similaridade com a sequência Pcarvalhoi\_clone17 recuperadas do genoma de *P. carvalhoi*.

**Figura 3:** Estrutura molecular predita para as cópias onde foram recuperadas potenciais TIRs

**Figura 4.** Histograma com os resultados do blast da open read frame (ORF) de hAT\_10 no genoma de *Pipa carvalhoi*. É possível observar que a maior densidade de matches estão em contigs menores de 5kb, marcados como blocos em preto em nosso esquema.

**Figura 5.** Histograma com os resultados do blast do open read frame (ORF) de hAT\_10 no genoma de *Xenopus tropicalis*. É possível observar que a maior densidade dos matches estão em contigs com menos de 2kb.

**Figura 6.** Gráfico obtido pela ferramenta DNAPipeTE com o conteúdo de elementos transponíveis no genoma de *Pipa carvalhoi*. Os resultados demonstram que cerca de 10% do genoma é composto por transposons de DNA, classe que inclui o elemento hAT.

**Figura 7:** Análise da distribuição dos fragmentos encontrados através do blast do elemento hAT\_10 no genoma de (A) *Xenopus tropicalis* (cópia consenso REPBASE) e (B) *Pipa carvalhoi* (contig pipa481784). A área pontilhada (vermelho) representa unidades repetitivas observadas dentro do elemento. Retângulo representa posição que a sequência Pcarvalhoi\_clone17 alinha na respectiva cópia.

**Figura 8.** Análise de cópias e cobertura do consenso do hAt\_10 contra o genoma de *Xenopus tropicalis*, evidenciando a fragmentação e divergência das cópias em relação ao consenso do elemento transponível.

**Figura 9.** Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Xenopus tropicalis*, onde é possível perceber um grande número de cópias repetitivas na região entre 4000-6000 pares de base, bem como na porção de 10.000 pares de base.

**Figura 10.** Análise de cópias e cobertura do consenso do hAT\_10 contra o genoma de *Pipa carvalhoi*, onde é possível observar que a maior parte das cópias deste elemento estão fragmentadas, mas uma delas, evidenciada pela linha em vermelho, abrange toda a região da ORF do hAT. A ausência de cobertura na região por volta de 4,1 kb parece indicar o fim da cópia do elemento de transposição no contig pipa481784.

**Figura 11.** Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Pipa carvalhoi*. É possível observar que a cobertura das repetições do início do contig é bem menor quando comparada a região de 4kb do elemento.

**Figura 12.** Análise de cópias e cobertura do consenso do hAT\_10 contra o genoma de *Pipa carvalhoi*. No consenso fica evidenciada a resolução da região de repetições do início da sequência, além da presença de mais repetições na região de 4kb.

**Figura 13.** Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Pipa carvalhoi*. A presença de potenciais MITES está evidenciada pelas setas verdes, presentes na região da ORF do elemento de hAT\_10.

**Figura 14.** Desenho esquemático da análise da ORF do elemento de hAT através da ferramenta ORFfinder, destacando que a sequência se trata de um hAT\_10 incompleto, uma vez que domínio Dimer ainda não está na ORF, ainda que o domínio de dedo de zinco (Zf-BED) esteja dentro da ORF.

**Figura 15.** Desenho esquemático da análise da ORF do elemento de hAT através da ferramenta ORFfinder, destacando que a sequência se trata de um hAT\_10, destacando a obtenção da ORF completa.

**Figura 16.** Desenho esquemático da análise da ORF consenso do elemento de hAT através da ferramenta ORFfinder. Com esta montagem, a ORF consenso apresentou um tamanho aproximado de 3,2 kb, sendo próxima ao tamanho da ORF original de *Xenopus tropicalis*.

**Figura 17.** Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Pipa carvalhoi*. É possível observar que, com a sequência consenso de hAT, as repetições encontradas na região da ORF do elemento, bem como na região de 4kb, permanecem sendo mapeadas no genoma da espécie.

**Figura 18.** Análise de cópias e cobertura do possível MITE 1. As primeiras e as últimas 100 bases são da região flanqueadora do potencial MITE. As linhas em vermelho representam os fragmentos que alinharam com esta região.

**Figura 19:** Análise de cópias e cobertura do possível MITE 2. As primeiras e as últimas 50 bases são da região flanqueadora do potencial MITE. As linhas em vermelho representam os fragmentos que alinharam com esta região.

## LISTA DE TABELAS

### **CAPÍTULO 1: Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome**

**Table 1:** Recorded cases of transposable elements giving rise to satellite, mini- and microsatellite DNA (satDNA), according to their mobility and the regions of the transposable elements that have given rise to the tandem repeat sequences. **ORF** (Open Reading Frame); **TR** (Tandem Repeat); **UTR** (Untranslated Region); **LTR** (Long Terminal Repeat), **TIR** (Terminal Inverted Repeats), **IR** (Inverted Repeat).

### **CAPÍTULO 2: Elementos de transposição como reservatórios de unidades repetitivas durante o surgimento de um novo DNA satélite: o caso dos elementos hAT no genoma de *P. carvalhoi* e de *X. tropicalis***

**Tabela 1:** Análise comparativa da identidade do clone isolado do genoma de *P. carvalhoi* (Pcarvalhoi\_clone17) em relação aos contigs recuperados do genoma da referida espécie (colunas 1 e 2) e sua comparação quanto à similaridade (%) do melhor hit recuperado no REPBASE. Valores de identidade (%) foram recuperados de comparações realizadas no BLAST. A análise de presença e tamanho (pb) de repetições invertidas terminais (TIRs) também é fornecida. *XT*: *Xenopus tropicalis*; (-): ausência de TIRs.

**Tabela 2:** Tabela com resultados das análises de blastn com base no consenso final de hAT\_10 gerado e os genomas de anuros já sequenciados, com foto no número de cópias desta repetição encontrados em cada genoma.

**Tabela 3:** Resultados obtidos pelo blastn comparando a sequência consenso obtida de hAT\_10 de *Pipa carvalhoi* com os genomas de anuros já sequenciados, buscando evidenciar as regiões de início e fim das repetições, destacando a cobertura em relação à query.

## LISTA DE ABREVIATURAS OU SIGLAS

2n - número diplóide  
BLAST – *basic local alignment search tool*  
CENP A – *centromere protein*  
DBS- *double – strand breaks*  
DNA – ácido desoxirribonucleico  
ENV – *envelope protein*  
FISH – Híbridaç o *in situ* fluorescente  
GAG – *GAG domain*  
H3 – histona H3  
IR – repetiç es invertidas  
LINE – *long interspersed nuclear element*  
LTR- longas repetiç es terminais  
NCBI – *National Center for biotechnology information*  
NAHR - *non-allelic homologous recombination*  
NHEJ - *non-homologous end joining*  
ORF- *open reading frame*  
pb – pares de base  
PCR – *polymerase chain reaction*  
SatDNA - DNA sat elite  
SINE – *short interspersed nuclear element*  
SNPs – *single nucleotide polymorphism*  
TE – elemento transpon ivel  
TIR – repetiç es terminais invertidas  
TDS – *target site duplication*  
UTR– *untranslated Region*



## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>18</b>
1.1. Reescrevendo o genoma: elementos transponíveis e a origem de novas sequências repetitivas... ..	18
1.2. Explorando as sequências genéticas móveis: elementos hAT em Anuros.....	22
1.3. Citogenética dos anfíbios pipídeos: estudos cromossômicos na família Pipidae .....	24
<b>2. OBJETIVOS.....</b>	<b>27</b>
2.1. Objetivo Geral .....	27
2.1.1. Objetivos Específicos.....	27
<b>3. RESULTADOS .....</b>	<b>28</b>
3.1. Capítulo 1: Transposable elements as a source of novel repetitive DNA in the eukaryote genome... ..	30
3.2. Capítulo 2: Elementos de transposição como reservatórios de unidades repetitivas durante o surgimento de um novo DNA satélite: o caso dos elementos hAT no genoma de <i>P. carvalhoi</i> e de <i>X. tropicalis</i> .....	48
3.3. Capítulo 3: Do the insertion preferences of transposable elements represent a strategy for the mitigation of genetic conflicts with the host? .....	85
<b>4. CONSIDERAÇÕES FINAIS.....</b>	<b>89</b>
<b>REFERÊNCIAS. ....</b>	<b>90</b>

## 1. INTRODUÇÃO

### *1.1 Reescrevendo o genoma: elementos transponíveis e a origem de novas sequências repetitivas*

No âmbito das sequências de DNA repetitivo, os elementos transponíveis (TE) representam uma parcela substancial do conteúdo genômico total dos eucariotos, desempenhando um papel fundamental nos variados processos biológicos dos genomas em que estão incorporados (Hayward, Gilbert, 2022), incluindo impacto na arquitetura genômica, na regulação gênica e na variabilidade das espécies.

Esses componentes são categorizados em duas classes de acordo com o sistema proposto por Wicker (2007): os *retrotransposons* (Classe I) e os *transposons de DNA* (Classe II), onde os elementos pertencentes à Classe I realizam sua transposição pelo genoma utilizando o mecanismo "copia e cola", o qual exige a participação de uma molécula intermediária de RNA. Em contrapartida, os transposons de DNA executam sua mobilidade sem essa intermediação, adotando o mecanismo "recorta e cola". Os elementos englobados dentro da classe I, ainda podem ser subdivididos do ponto de vista estrutural em elementos que apresentam longas repetições terminais (*long terminal repeat* - LTRs) e os que não possuem essas repetições, nomeados de não-LTR (Wicker et al., 2007; Kapitonov e Jurka, 2008).

Devido à relevância da mobilidade, os elementos móveis são ainda separados em *autônomos*, quando todos os domínios necessários para efetuar uma transposição estão presentes e conservados; e *não-autônomos* quando possuem uma região codificante degenerada, dependendo de outros elementos para se moverem no genoma (Wicker, 2007).

Os TEs podem afetar os genomas através dos eventos de mobilização, com impactos mais locais variando de acordo com a extensão e localização do sítio de inserção ou pós-inserção, influenciando de maneira importante a estrutura do genoma e a dinâmica dos cromossomos (McClintock, 1950; Kazazian et al., 1988; Deininger and Batzer 1999; Johnson et al. 2013; Klein e O'Neil, 2018).

Um dos importantes impactos se refere à interferência na expressão gênica, podendo induzir aumento ou diminuição da expressão de genes adjacentes, com inserções ocorrendo em locais como íntrons, éxons e regiões promotoras (Gill et al., 2021). Em *Drosophila*, por exemplo, foram identificados mais de quatrocentos TEs associados à variação de expressão gênica em genes próximos, apresentando tanto regulação positiva quanto negativa (Rech et al., 2022).

Além disso, essas repetições podem desempenhar um papel importante na estrutura cromossômica contribuindo para o alongamento e manutenção dos telômeros (Kordyukova et al., 2018), que são essenciais para a estabilidade e replicação cromossômica; bem como influenciando a estrutura e a função dos centrômeros, responsáveis pela segregação cromossômica adequada durante a divisão celular (Gržan et al., 2020; Talbert e Henikoff, 2020; Amorim et al., 2021). Esses papéis estruturais dos TEs contribuem para o tamanho e a organização do genoma, impactando a arquitetura geral dos genomas, facilitando a inovação e a diversificação genética.

Os genomas eucariotos, por exemplo, variam notoriamente em relação ao seu tamanho, e uma das causas dessa variação está ligada à quantidade de elementos transponíveis (Elliot e Gregory, 2015). Em estudos com tamanhos de genoma em Caudata realizados em seis espécies de salamandras, foi observado que entre 25% e 47% dos genomas eram compostos por sequências repetitivas, predominando a presença de retrotransposons – LTR (Metcalf e Cesane, 2013).

Do ponto de vista estrutural, existem evidências de rearranjos cromossômicos como inserções, deleções, inversões e duplicações mediados por elementos transponíveis, em praticamente todas as linhagens de vertebrados, podendo estar envolvidos até mesmo em processos de especiação (Bohne et al., 2008). Estes eventos podem ocorrer de maneira indireta, por meio da recombinação homóloga, ou diretamente, quando a quebra da dupla-fita causada pelo processo de transposição é corrigida de maneira errônea pelo sistema de reparo do DNA, utilizando sequências homólogas ectópicas como modelo para tal (Gray, 2000; Hedges e Deininger, 2007).

Para melhor compreender o papel e a dinâmica dos elementos transponíveis, é essencial conhecer o seu ciclo de vida característico, indo de sua inserção até a chamada “eliminação”. Em um primeiro momento ocorre a fase de invasão do TE, quando este sofre uma duplicação e se insere em um novo locus cromossômico, seguido por um aumento em seu número de cópias (fase de amplificação), e maturação, onde os mecanismos de defesa do hospedeiro entram em ação para silenciá-los e/ou inativá-los (Paço et al., 2019; Zattera e Bruschi, 2023).

Uma etapa especialmente intrigante dentro desse ciclo é sua última etapa: o estágio de "degeneração" ou “morte”. Após a atuação dos mecanismos de silenciamento, promovido por mecanismos de regulação presentes no genoma hospedeiro, há a possibilidade de inativação dessas sequências, resultado da degeneração das regiões contendo os domínios conservados por meio do acúmulo de mutações deletérias, interferindo em seu potencial de mobilização (Brookfield, 2005; Sigman et al., 2016). Quando estes elementos são inativados o TE pode seguir diversos caminhos como: se manter inativo, sofrer processos de domesticação molecular ou, até mesmo, dar origem a novas sequências repetitivas como *miniature inverted-repeat transposable elements* (MITEs), DNA satélites e outras sequências organizadas em tandem (Paço et al., 2019; Zattera e Bruschi, 2022).

MITEs são elementos transponíveis curtos e não-autônomos, apresentando entre 100 e 800 pares de base (pb), derivados de transposons de DNA (Xin et al., 2019; Castanera et al., 2021). Tais elementos apresentam predileção por inserção em regiões ricas em genes (Venkatesh e Nandini, 2020), e apresentam um par de repetições terminais invertidas (TIR) conservadas ( $\geq 10$  pb), que desempenham o papel de locais de ligação para a transposase, juntamente com um par de duplicações curtas nos sítios-alvo (TSDs) em média de 2 a 10 pb, situados na extremidade exterior de cada TIR (Feschotte et al., 2002; Ye et al., 2016; Venkatesh e Nandini, 2020).

As sequências repetidas em tandem, por sua vez, estão extensivamente presentes em organismos procariotos e eucariotos, organizadas em uma orientação conhecida

como “cabeça-cauda”, onde a sequência é repetida uma após a outra, em uma espécie de arranjo linear (Legendre et al., 2007; Zhao et al., 2014). Dentro dessas sequências, são encontrados os DNA satélites, classificados de acordo com o tamanho do seu motivo repetitivo em pares de base como microssatélites [1 a 10 pb], minisatélites [10 a 100 pb], e satélites [repetições maiores de 100 pb] (Thakur et al., 2021; Zhu et al., 2021).

Existe uma relação bastante complexa entre TE e DNA satélites (satDNA) que pode ter efeitos cruciais na estrutura e evolução genômica (Belyayev et al., 2020). Elementos de transposição já foram reportados dando origem à satDNA em diversos grupos como em roedores do gênero *Ctenomys* (Rossi et al., 1993), em plantas da espécie *Arabidopsis thaliana* (Kapitov e Jurka, 2000), em galinhas (Shang et al., 2010), em milho (Sharma et al., 2013), entre outros.

Quando consideramos o grupo de anuros, casos como este já foram reportados em *Xenopus leavis* (Pasero et al., 1993; Hikosaka et al., 2004), onde elementos do tipo MITE e SINE deram origem aos DNA satélites Xstir e Satellite 1, respectivamente, e na espécie de *Xenopus tropicalis*, com o elemento MITE de TC1-mariner dando origem ao satDNA miDNA4 (Scalvenzi e Pollet, 2014).

DNAs satélite (satDNAs), que têm origem em elementos transponíveis (TEs), frequentemente resultam da amplificação em arranjos em tandem de partes estruturais desses elementos móveis ainda que, alternativamente, essas sequências possam surgir a partir de elementos completos (como retrotransposons) envolvendo uma combinação de elementos móveis diversos que sofreram truncamento (Mestrovic et al., 2015).

Pesquisas conduzidas por McGurk e Barbash (2018) em *Drosophila melanogaster* revelaram padrões distintos na origem de repetições em tandem a partir de elementos transponíveis. Conforme indicado pelos autores, os retrotransposons-LTR apresentam uma maior ocorrência de recombinação ectópica entre suas repetições terminais, ao passo que os retrotransposons não-LTR e os transposons de DNA demonstraram maior probabilidade de originar outras repetições por meio de múltiplas inserções no mesmo locus.

Em seu processo de mobilidade, os TEs frequentemente causam quebras na dupla fita de DNA [*double-strand break* - DSB] e muitos dos eventos de instabilidade no genoma envolvendo estes elementos, são resultados de processos de reparo errôneos destas quebras (Hedges e Deininger, 2007). Existem diversos mecanismos envolvidos no reparo, sendo, em grande maioria, dependentes da homologia para que a informação contida na sequência seja mantida.

A união de extremidade não-homóloga (*non-homologous end joining* - NHEJ) e a recombinação homóloga não-alélica (*non-allelic homologous recombination* - NAHR) parecem ser mecanismos comuns para o reparo das quebras de dupla fita, sendo que eventos como deleções e/ou duplicações podem ocorrer entre os TEs (Davis and Chen 2013; Gu et al. 2008; Robberecht et al 2013; Kim et al., 2019).

A mobilização culmina na formação de “sítios alvo de duplicação” (*target site duplication* - TSD) no ponto de inserção do elemento (Lavie et al., 2004), que podem promover a homologia necessária para que o mecanismo de NHEJ possa restaurar esta quebra e o posterior deslizamento da polimerase pode ocasionar a expansão do número de cópias do elemento (Gilbert et al., 2015; Zattera e Bruschi, 2022).

Com o exposto acima, é possível observar inúmeras evidências do envolvimento de elementos transponíveis (TEs) na remodelação do genoma, atuando como base para o aparecimento de elementos repetitivos adicionais, participando como reguladores da expressão genética ou sendo cooptados para novas funções dentro do genoma. Isto destaca a adaptabilidade destes elementos dentro dos genomas de organismos eucarióticos, bem como a importância de sua compreensão nos genomas onde estão inseridos.

### ***1.2 Explorando as sequências genéticas móveis: elementos hAT em anuros***

Seguindo a classificação de Wicker *et al.* (2007), os elementos de classe II podem ser subdivididos em duas subclasses, diferenciadas principalmente pelo número de fitas de DNA que são clivadas durante o mecanismo de transposição. A subclasse I é caracterizada pelo mecanismo envolvendo a clivagem de ambas as fitas de DNA, sendo

composta por duas ordens: a TIR (*Terminal inverted repeat*), representada por nove superfamílias (*PiggyBac*, *Tc-1-Mariner*, *Transib*, *CACTA*, *hAT*, *Merlin*, *Mutator*, *P* e *PIF-Harbinger*) e a ordem Crypton, que apresenta apenas a uma superfamília que dá nome à esta ordem (Wicker *et al.*, 2007). A subclasse II apresenta as ordens Helitron e Maverick, que abrangem superfamílias que também dão nomes às suas respectivas ordens e que são caracterizadas pela clivagem envolvendo apenas uma das fitas de DNA em seu processo de transposição (Wicker *et al.*, 2007). Estando presente dentro da subclasse I, a superfamília de *hAT* recebe este nome devido à três transposons de DNA, o (i) *hobo*; (ii) Ativador (*Ac*); e (iii) *Tam3* (Rubin *et al.*, 2001), e os elementos presentes em *hAT* apresentam algumas particularidades relacionadas à sua estrutura. Tais elementos apresentam tamanho menor que 4kb, exceto elemento *Gulliver* relatado em *Chlamydomonas reinhardtii* (Ferris, 1989), TSDs (*target site duplication*) com oito nucleotídeos, além de uma repetição terminal invertida curta (TIR) com um tamanho variável entre 5-27 pb (Wicker *et al.*, 2007).

Considerando que a origem destes elementos pode ser datada de antes da separação entre plantas, fungos e animais, elementos da superfamília de *hAT* apresentam uma ampla distribuição entre os eucariotos (Wicker *et al.*, 2007). E tais transposons já foram reportados em diferentes espécies de anfíbios.

Em *Xenopus tropicalis*, referência em estudos genômicos de anfíbios já foi relatada forte presença de transposons de DNA, principalmente das superfamílias *Tc-Mariner*, *Harbinger*, *PiggyBac* e *T2/Kolobok* e *hAT* (Hellsten *et al.*, 2010; Chalopin *et al.*, 2015; Sun *et al.*, 2015). Ainda nos estudos conduzidos por Sun *et al.* (2015) através do sequenciamento completo do genoma de *Nanorana parkeri*, também foi constatado a presença do elemento hAT-Ac na espécie. No rascunho da montagem de genoma da espécie *Rhinella marina* foi possível encontrar grande presença de transposons de DNA, sendo o elemento hAT-Ac o mais abundante dentro desta classe (Edwards *et al.*, 2018).

Porém, ainda que estes elementos tenham sido reportados dentro das espécies de anfíbios, o grupo ainda carece de uma análise mais aprofundada acerca destes

transposons, bem como o papel que tais sequências podem desempenhar dentro das regiões repetitivas dos genomas onde estão inseridos.

### ***1.3 Citogenética dos anfíbios pipídeos: estudos cromossômicos na família***

#### ***Pipidae***

A classe Amphibia apresenta uma notável diversidade, abrigando atualmente um total de 8.674 espécies, que são divididas em três ordens: Caudata (822 sp.), Gymnophiona (221 sp.) e Anura, sendo essa a mais numerosa, com um total de 7.631 espécies representantes (Frost, 2023).

No grupo dos anuros, merece destaque a família Pipidae, composta por duas subfamílias: Dactylethrinae (34 sp.), que engloba os gêneros *Hymenochirus* (4 sp.), *Pseudhymenochirus* (1 sp.) e *Xenopus* (29 sp.); e Pipinae, que tem como seu único representante o gênero *Pipa* (7 sp.) (Frost, 2023).

Pipidae apresenta uma ampla distribuição, abrangendo tanto a África quanto a América, além de ser reconhecida como uma das linhagens mais antigas entre os anuros, constituindo um grupo monofilético (Cannatella; Trueb, 1988; Irisarri et al., 2011; Frost, 2023). Acredita-se que sua diversificação teve início a partir de um ancestral comum, em um evento vicariante durante a separação da Pangeia, onde o isolamento subsequente desencadeou a diversificação da linhagem que deu origem ao gênero *Pipa*, o único gênero não africano dentro dessa família (Trueb et al., 2005; Irisarri et al., 2011).

Dentro deste grupo taxonômico, o gênero *Xenopus* se destaca como um modelo biológico amplamente reconhecido e utilizado em diversas áreas de pesquisa (Lillo et al., 2011, e Uno et al., 2013). Este grupo apresenta como característica marcante, a presença de poliploidia, onde um organismo possui mais de dois conjuntos completos de cromossomos, um processo que provavelmente aconteceu devido a um evento de aloploidização (Schmid et al., 2015). Por exemplo, *Xenopus laevis* e *X. borealis* são espécies aloploidoides derivadas de um ancestral com  $2n = 18$  cujo genoma inteiro foi



duplicado aproximadamente há aproximadamente 18 milhões de anos, apresentando como resultado em um genoma tetraploide de 36 cromossomos (Session et al., 2016).

Por outro lado, a espécie de *Xenopus tropicalis* destaca-se como uma espécie diplóide ( $2n=20$ ) (Tymowska et al., 1973; Mezzasalma et al., 2015), além de ser a pioneira entre os anuros a ter seu genoma completamente sequenciado (Hellsten et al., 2010), em um projeto realizado pelo Joint Genome Institute (JGI) no ano de 2001.

Dentro do gênero *Pipa*, estão presentes as espécies *P. pipa*, *P. carvalhoi*, *P. parva*, *P. myersi*, *P. snethlageae*, *P. arrabali* e *P. aspera*, sendo estas encontradas nas regiões norte da América do Sul e Panamá (Frost, 2023). Os membros deste gênero são exclusivamente aquáticos, e possuem corpos achatados dorso-ventralmente bem como membros traseiros adaptados para a natação (Pough et al., 2001; Fernandes et al., 2011).

Do ponto de vista citogenético, os cariótipos conhecidos e descritos para este grupo, permitem observar uma discrepância de números diplóides, onde *P. parva* apresenta  $2n=30$  (Morescalchi, 1968), *P. carvalhoi* possui  $2n=20$  (Mezassalma et al., 2015; Zattera et al., 2019) e *P. pipa* possui  $2n=22$  cromossomos (Wickbom, 1950; Morescalchi, 1968).

Estudos conduzidos por Mezzasalma et al (2015), buscaram compreender qual seria a condição plesiomórfica dentro de Pipidae. Utilizando cariótipo de *Rhinophrynus dorsalis*, táxon irmão de Pipidae com  $2n=22$ , os autores inferiram os possíveis rearranjos envolvidos na transição de seu cariótipo ( $2n = 22$ ) para os de  $2n = 20$  encontrados em *P. carvalhoi* e *X. tropicalis*, destacando que o ancestral de Pipidae teria  $2n=20$ , sendo esta a via mais parcimoniosa dentro da filogenia, exigindo menos rearranjos ou passos evolutivos para a transição do cariótipo.

Considerando exclusivamente o gênero *Pipa*, este também parece representar o caminho mais parcimonioso dentro da filogenia. Isso porque, para as espécies do gênero evoluírem de um ancestral com  $2n=30$  para o cariótipo observado em *P. carvalhoi*, seriam necessários mais eventos evolutivos, como numerosas fusões, por exemplo (Mezzasalma et al., 2015; Zattera et al., 2019).

Em estudos comparativos entre as espécies *P. carvalhoi* e *X. tropicalis* utilizando sequências altamente conservadas como histona H3, foi possível verificar homologias entre as duas espécies, com clusters de histona estando presente nos cromossomos 5 e 6 em *P. carvalhoi*, *X. tropicalis* e *X. laevis*, e nos pares 8 e 9 somente nos cromossomos de *P. carvalhoi* e *X. tropicalis* (Zattera et al., 2020).

Entretanto, ao considerar os repetitivos de DNA mais variáveis, como sequências de microssatélites, foram observados padrões semelhantes entre as espécies, como (CAG)<sub>10</sub>, (CGC)<sub>10</sub> e (GATA)<sub>10</sub>, mapeados nas regiões terminal/subterminal dos cromossomos; mas também foram identificadas divergências nos microssatélites de (CA)<sub>15</sub> e (GA)<sub>15</sub>, presentes nas regiões terminais e subterminais em *X. tropicalis*, e nas regiões centroméricas e pericentroméricas de *P. carvalhoi* (Zattera et al., 2020).

Uma vez estabelecida a proximidade filogenética e citogenética entre os gêneros de *Pipa* e *Xenopus*, sugerimos a adoção de *Pipa* como grupo central em nossas investigações, buscando aproveitar os conhecimentos gerados acerca do gênero *Xenopus*, a fim de elucidar a dinâmica dos elementos repetitivos de DNA no genoma do gênero *Pipa*.

## 2. OBJETIVOS

### 2.1. OBJETIVO GERAL

Explorar o processo de surgimento, manutenção e evolução de sequências de DNAs repetitivos organizados “*in tandem*” a partir de elementos genéticos móveis e avaliar no genoma de *P. carvalhoi* e *X. tropicalis* a relação entre elementos de transposição da família hAT e novos DNAs satélites

#### 2.1.1 OBJETIVOS ESPECÍFICOS:

- Buscar evidências de processo de surgimento de DNAs satélites a partir das sequências de transposons da família hAT;
- Avaliar o grau de repetibilidade das sequências-alvo nos genomas de *P. carvalhoi* e *X. tropicalis*;
- Analisar e caracterizar a organização molecular das sequências repetitivas encontradas neste estudo;
- Percorrer estratégias metodológicas que permitam caracterizar elementos genéticos móveis portadores de matrizes de elementos repetitivos internos;
- Avaliar criticamente o papel das inserções de elementos de transposição em sítios genômicos preferenciais, confrontando dados da literatura.

### 3. RESULTADOS

Os resultados obtidos foram organizados na forma de três capítulos, conforme descrito a seguir:

**Capítulo 1: “Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome”** – Artigo científico publicado na revista *Cells* (DOI: 10.3390/cells11213373) que explora dados da literatura para avaliar casos de surgimento de novos elementos repetitivos a partir de elementos genéticos móveis em genomas de organismos eucariotos. São exploradas questões como a estrutura molecular do elemento, sua organização cromossômica, o papel do “ciclo de vida” dos TEs neste surgimento e os mecanismos genéticos que podem estar envolvidos. Este artigo apresenta a sistematização dos dados de interconversão entre elementos repetitivos nos genomas e destaca a perspectiva de novas questões que ainda devem ser mais exploradas em estudos de evolução genômica.

**Capítulo 2: Elementos de transposição como reservatórios de unidades repetitivas durante o surgimento de um novo DNA satélite: O caso do elemento hAT no genoma de *P. carvalhoi* e *X. tropicalis*”**. Neste capítulo apresentamos dados experimentais e de análises de bioinformática para a caracterização de elementos repetitivos originados de transposons da família hAT no genoma de *P. carvalhoi* e *X. tropicalis*, enfrentando os desafios metodológicos de avaliar matrizes de unidades monoméricas presentes na estrutura molecular de TEs.

**Capítulo 3: “Do the insertion preferences of transposable elements represent a strategy for the mitigation of genetic conflicts with the host?”** – Manuscrito submetido para publicação na revista *Evolution* que discute as implicações da preferência de inserção de TEs no genoma hospedeiro para seu sucesso na invasão de um novo genoma. O manuscrito foi descrito como forma de posicionar os achados de

simulações computacionais de Munasinghe et al., (2023) frente aos dados experimentais disponíveis na literatura. O trabalho foi submetido na sessão “Digest” do periódico *Evolution*, destinada a “digerir”, simplificando artigos com grande relevância na área demonstrando sua aplicabilidade no conhecimento atual disponível na literatura.

### **3.1. CAPÍTULO 1:**

## **TRANSPOSABLE ELEMENTS AS A SOURCE OF NOVEL REPETITIVE DNA IN THE EUKARYOTE GENOME**

Review

# Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome

Michelle Louise Zattera <sup>1</sup>  and Daniel Pacheco Bruschi <sup>2,\*</sup> 

<sup>1</sup> Departamento de Genética, Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba 81530-000, PR, Brazil

<sup>2</sup> Departamento de Genética, Laboratório de Citogenética Evolutiva e Conservação Animal, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba 81530-000, PR, Brazil

\* Correspondence: danielbruschi@ufpr.br

**Abstract:** The impact of transposable elements (TEs) on the evolution of the eukaryote genome has been observed in a number of biological processes, such as the recruitment of the host's gene expression network or the rearrangement of genome structure. However, TEs may also provide a substrate for the emergence of novel repetitive elements, which contribute to the generation of new genomic components during the course of the evolutionary process. In this review, we examine published descriptions of TEs that give rise to tandem sequences in an attempt to comprehend the relationship between TEs and the emergence of de novo satellite DNA families in eukaryotic organisms. We evaluated the intragenomic behavior of the TEs, the role of their molecular structure, and the chromosomal distribution of the paralogous copies that generate arrays of repeats as a substrate for the emergence of new repetitive elements in the genome. We highlight the involvement and importance of TEs in the eukaryote genome and its remodeling processes.

**Keywords:** satellite DNA; centromeric region; TE life cycle; tandem repeats



**Citation:** Zattera, M.L.; Bruschi, D.P. Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome. *Cells* **2022**, *11*, 3373. <https://doi.org/10.3390/cells11213373>

Academic Editor: Laura Fanti

Received: 29 July 2022

Accepted: 26 September 2022

Published: 26 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The duplication and mobilization cycles of transposable elements (TEs) are evolutionary processes that enrich the paralogous copies that constitute the repetitive DNA content of the eukaryote genome. This repetitive DNA can be subdivided into two categories: tandem repeat sequences, which represent copies organized in juxtaposition to one another (e.g., DNA satellites, minisatellites, and microsatellites), or dispersed sequences, which include the transposable elements themselves [1,2].

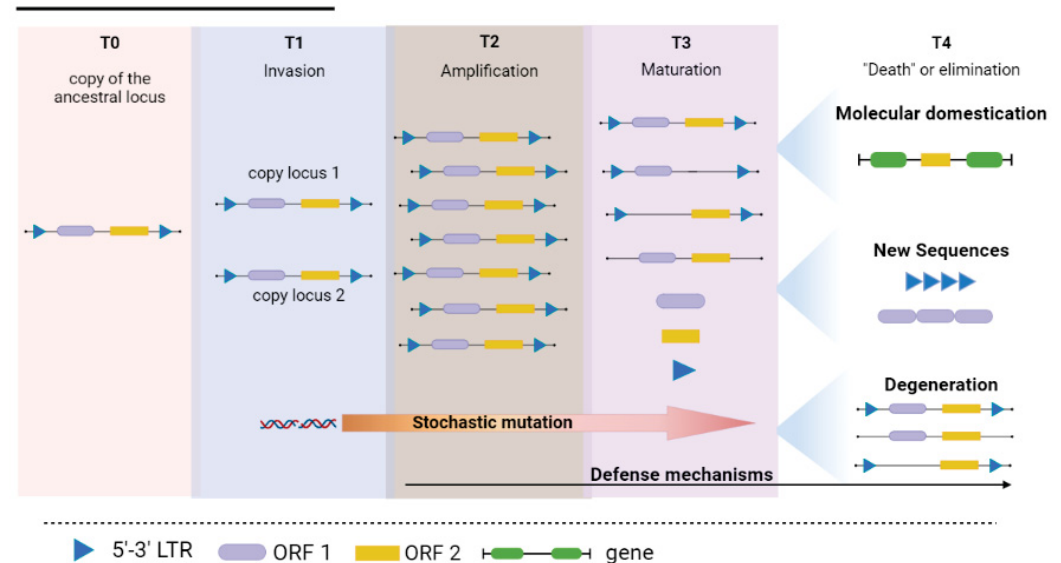
In most cases, transposable elements and tandem repeats are studied independently, but evidence has been found that TEs may be involved in the origin of a library of tandem repeats that is typically dispersed throughout the eukaryote genome, where it plays a fundamental evolutionary role [3,4]. The sequences of homologies of satDNAs and transposons or retrotransposons that have been identified in many species indicate the existence of an intimate evolutionary relationship between the TEs and the emergence of tandem repeat sequences, which implies that the TEs are involved in the reshaping of the genomic architecture [3,5].

The conversion of one type of repetitive element into another has been reported in all the principal branches of the eukaryote tree of life [6–8]. Here, we review case studies of the emergence of repetitive DNA, focusing on the new satellite DNA (satDNA) families derived from TEs with the aim of understanding (i) the intragenomic behavior of the TEs, (ii) the role of their molecular structure, and (iii) the chromosomal distribution of the paralogous copies. We discuss the genetic mechanisms that produce the TE copies and contribute to the molecular co-option of these elements during the evolution of the genome and chromosomes, most of which are revealed by the interplay and turnover of the different repetitive classes in the genome.

There are two major categories of transposable elements: classes I and II. Class I elements, which are also known as retrotransposons, are dependent on the RNA for their transposition in the genome, while class II elements (DNA transposons) do not depend on any retrotranscription mechanism for their mobility [9].

The recruitment of copies of the TEs as a substrate for the evolutionary emergence of new tandem repeats and satDNA families is directly dependent on the behavior of these elements in the host genome. The intra-genomic behavior of the TEs in the host genome follows a four-phase life cycle (Figure 1), starting with (i) the birth or invasion phase, when the TEs are inserted into a new locus, followed by (ii) amplification, (iii) maturation, and finally, (iv) death or degeneration [10–12].

### Life cycle of a Transposable Element



**Figure 1.** Life cycle of a transposable element, using LTR transposons with their respective Open Reading Frames (ORFs) as an example. **T0**—simple copy of the element in the genome; **T1—Invasion:** the chromosomal locus acquires at least one repeat unit; **T2—Amplification:** an increase in the number of copies present in the genome, either by burst events (in the case of class I elements) or the repair or homologous recombination of double-stranded DNA (class II); **T3—Maturation:** the elements are inactivated or silenced through epigenetic silencing, piRNAs, DNA methylation or other mechanisms, such as mutations; **T4—Death or degeneration:** the elements may be either eliminated from the genome, undergo molecular domestication, begin to exercise new functions or give rise to new repetitive sequences, such as tandem repeats. **LTR** = Long Terminal Repeat; **ORF** = Open Reading Frame. (Source: the authors / created in [Biorender.com](https://www.biorender.com), accessed on 5 September 2022).

The birth or invasion phase (Figure 1—T1) occurs when a new chromosomal locus acquires at least one copy of a TE. In most cases, the TE is transmitted vertically through the evolutionary lineage by splitting from the ancestral species, while invasion occurs through new transposition events, either by the recombination of sequences already present in the genome or events of horizontal transfer [13,14].

As a new repeat unit of a class I element (retrotransposon) is established in the genome, the number of copies begins to increase rapidly (amplification phase; T3 in Figure 1). These paralogous copies, generated by retrotransposition, are distributed randomly throughout the genome [15,16]. Genomes of all characterized higher eukaryotes possess examples of transposable element (TE) bursts [17]. Burst events can cause a local accumulation of TEs in the chromosomes, generating extensive clusters of TEs [18–21].

Events of this type are more frequent in germ cells due to the temporary relaxation of the epigenetic control of the TEs during the early development of the germline. This opens a temporal window that may allow these elements to escape from their constraints and



propagate in the host genome [22,23]. Amplification occurs primarily in male germ cells due to the continuous spermatogenesis occurring throughout the life span of the animal, whereas female germlines are arrested in meiosis and do not undergo the temporary relaxation of epigenetic control observed in the male germline [24,25].

By contrast, the repair mechanism of double-strand (subclass I) or single-strand breaks (subclass II; see [9]) that occur following mobilization events is a central genetic process that results in the addition of a number of copies of the DNA transposons, given that the chromosomal gap generated by the excision of a transposon sequence is repaired by a homologous recombination mechanism, which results in the reintroduction of the transposon at the donor site [26]. If the mobilization event occurs during the S phase, the homologous recombination can use a sister chromatid as a template for the repair of the chromosome gap, which also restores the excised element and results in an increase in the number of copies of the class II elements [26,27].

The maturation phase (T4 in Figure 1) is the most prolonged stage of the process, and is the most subject to the action of mechanisms of molecular evolution. The host genome can evolve defenses to stop the spread of the TEs by temporary epigenetic inactivation, irrespective of the mobilization mechanism involved in the insertion of the new TE [28]. During this phase, the silencing of transposition activity and the relaxation of selection pressures on the paralogous copies contribute to the accumulation of random mutations in the molecular structure of the TE sequences [29,30].

One of these pathways is through the silencing of the RNA by small RNA molecules—endogenous siRNAs (endo-siRNAs) or PIWI-associated RNAs (piRNAs)—which guide the process by modulating chromatin states or targeting the degradation of the RNA [31]. The piRNA pathway suppresses transposon activity in the metazoan germline, whereas the endo-siRNAs repress TE activity in the somatic tissue [31,32].

The transposition of the TEs may also be suppressed by DNA methylation, in which a methyl group is added covalently to the C'5 position of the cytosine (5-methylcytosine) [31,33]. This methylation is associated with repression of the TE, as first observed in the retrotransposon Activator (Ac), Suppressor-mutator (Spm), and Mutator (Mu) of maize [34–36]. This epigenetic suppression by DNA methylation is thought to contribute not only to the silencing of transcription, but also to the formation of the heterochromatic regions of the genome [37,38]. Epigenetic histone modifications, such as acetylation, phosphorylation, and methylation, may also be involved in the repression of the TEs [39]. In the genus *Arabidopsis*, for example, the histone methylation of H3K9me2 and H3K27me1 appears to contribute to the transcriptional silencing of some TEs [40], while histone deacetylase triggers the transcriptional activity of the different TEs [41]. The long-time prevalence of TEs under epigenetic control results in the progressive accumulation of mutations (substitutions or indels), which leads, in turn, to the degeneration of the sequence and a loss of identity, which either reduces or eliminates its capacity for amplification in the genome (Figure 1—T3). Recombination is often suppressed in the heterochromatic regions [42,43], which appears to be one of the reasons why these portions of the chromosome are prone to the accumulation of TEs, maintained by unequal crossing-over or genetic drift [44,45].

Stochastic mutations that interfere with the mobilization capacity of the silenced TE sequences are either fixed or lost primarily through genetic drift, which may represent the degeneration or death phase of the respective locus (Figure 1—T4). In this case, the heterochromatin is often the “final resting place” of dysfunctional TEs, due to the low frequency of recombination in these regions [44,45], leading to the formation of a “TE cemetery” [45,46], where the TEs may play an important role in the structure of the heterochromatin itself [47]. Given the high level of degeneration of the TE sequences, the recognition of genomes by searching for homologies often recovers contigs that are not very similar to the active copies. This hampers the understanding of their evolutionary history in the host genome.

The insertion and accumulation of some TEs in the vicinity of genes, for example, or their degenerate copies, represent an opportunity for molecular evolution or domestica-

tion [48–50]. Numerous studies have revealed protein motifs or repetitive regions of TEs being recruited for new genomic functions, providing an important source of novel sequences for the evolution of the genome [51–53]. New regulatory sequences and new protein-coding or even non-coding RNAs may play a beneficial role in the host genome [11,54,55], and may also be involved in a number of different processes of the regulation of gene expression. These processes include the regulation of transcription by the domesticated LTR retrotransposons and the action of the Amniota SINE1 element, which enhances the genes encoding the fibroblast growth factor of the inhibitor of an apoptosis protein family [56,57].

Given this, the ability of the TEs to provide a substrate for evolutionary innovations has attracted considerable interest over the years, and has been the focus of numerous studies of eukaryote species, which have found evidence of their involvement in the evolution of the genome through their reorganization.

## 2. Transposons as a Source of Repetitive Units for the Emergence of Tandem Repeats

A number of studies have provided examples of TE sequences that give rise to new repetitive classes, such as microsatellites, minisatellites, and satellite DNA [3,5,11,54,58]. Transposable elements represent an important substrate for genomic remodeling, and the emergence of new repetitive sequences, which generate a library of short repeat arrays that will subsequently be dispersed through the genome, to eventually become novel tandem repeats [59] (Figure 1—T4).

Transposable elements may contain sites predisposed for the formation of microsatellite DNA, which favors the dispersal of these repetitive units in the genome [5,58,60]. In the human genome, for example, approximately 23% of all tandem arrays (satellite, mini- and microsatellite sequences) are derived from TEs [61].

### 2.1. Classes and Mobility of the TEs

Studies of a range of different organisms have revealed a subtle prevalence of class I elements as a source of repetitions in the genome. Retrotransposons of the SINE (Short Interspersed Nuclear Element) superfamily, in particular the Alu elements, are a compelling example of this process. The non-autonomous SINEs, which range in size from 100 to 600 base pairs (bps), are widely distributed in the eukaryotes, where they play an important role in the organization of the genome, given their involvement in cell survival during many different types of physiological stress, for example [62]. The Alu elements, which are primate-specific, are considered to be the most widespread of the transposable elements, representing approximately 11% of the human genome [63–65]. These elements are considered to be the origin of the p $\lambda$ g3 satDNA and (GAA) $n$ -type repeats, which are both found in the human genome [66–68], and the A-rich microsatellites in the primates [69]. Approximately 7276 minisatellites in the human genome have been derived from transposable elements, and 2663 are associated with Alu elements [61].

Non-autonomous TEs prevail as a source of the sequences for the formation of novel repetitive elements (Table 1), and the MITE (Miniature Transposable Elements with inverted repetitions) elements may be an especially interesting group for the understanding of this process. The MITE-type elements are non-autonomous, class II elements of approximately 400 bps, which are characterized by inverted repetitions flanking a variable region [9,70,71]. The MITE repeats are known to give rise to (GTCY) $n$  repeats in lepidopterans [72], Xstir satDNA in *Xenopus leavis* [73], miDNA4 in *Xenopus tropicalis* [74], D1100 satDNA in rye [75], a number of different types of satellite DNA in *Messor* ants [76], and HindIII satDNA in bivalves [77].

**Table 1.** Recorded cases of transposable elements giving rise to satellite, mini- and microsatellite DNA (satDNA), according to their mobility and the regions of the transposable elements that have given rise to the tandem repeat sequences. **ORF** (Open Reading Frame); **TR** (Tandem Repeat); **UTR** (Untranslated Region); **LTR** (Long Terminal Repeat), **TIR** (Terminal Inverted Repeats), **IR** (Inverted Repeat).

Species	Type of Transposable Element (TE)	Superfamily	Class	Mobility	Region of the TE	New Sequence	Reference
<i>Pan paniscus</i> and <i>Hylobates lar</i>	Alu elements	SINE Alu Family	Class I	non-autonomous	3' oligo(dA) tail and A-rich middle region	A-rich primates' microsatellites	[69]
<i>Homo sapiens</i>	Alu   LTR – retrotransposons	SINE Alu Family	Class I	non-autonomous	TR may occur in any region of the TE	7276 Minisatellites	[61]
<i>Homo sapiens</i>	Alu	SINE Alu Family	Class I	non-autonomous	-	pAg3, pMSL, pMS43, and pMS228	[66]
<i>Homo sapiens</i>	Alu	SINE Alu Family	Class I	non-autonomous	Near to 3' -UTR.	Three minisatellites	[68]
<i>Homo sapiens</i>	Alu	SINE Alu Family	Class I	non-autonomous	3' oligo(dA) tail.	(GAA) <sub>n</sub>	[67]
Mouse genome	SINE B1	SINE Superfamily	Class I	non-autonomous	GAGGCA dimmer within the SINE	(GGCAGA) <sub>n</sub>	[78]
Mouse genome	MaLR	Retrotransposon-like superfamily	Class I	non-autonomous	LTR	Ms6-hm e Hm-2	[79]
<i>Ctenomys</i> sp.	retroviral genome	-	Class I	-	LTR	RPCS satDNA	[80]
<i>Phodopus roborovskii</i>	LINE-1 elements	LINE	Class I	autonomous	ORF2	PROsat	[81]
Dolphin	LINE-1	LINE	Class I	autonomous	-3' UTR	Common Cetacean Satellite	[82]
<i>Phoca vitulina concolour</i>	LINE-1	LINE	Class I	autonomous	ORF2	Pvc 20	[83]
Genus <i>Messor</i>	Mariner-like ( <i>Mboumar</i> )	Tc1/mariner	Class II	autonomous	Mariner is found inside the satDNA	satDNA	[84]
<i>Gallus gallus</i>	CR1	CR1 family of LINES	Class I	autonomous	-3' UTR and a partial coding region of ORF 2	Hinfl (SCR1)	[7]
<i>Helicoverpa zea</i>	HsSINE1 MITE-like	SINE Superfamily	Class II	non-autonomous	5'-IR	(GTCY) <sub>n</sub>	[72]

Table 1. Cont.

Species	Type of Transposable Element (TE)	Superfamily	Class	Mobility	Region of the TE	New Sequence	Reference
<i>Drosophila virilis</i> , <i>Drosophila americana</i> , and <i>Drosophila biarmipes</i> .	DINEs	Helitron	Class II	non-autonomous	Central tandem repeats (CTRs)	satDNA arrays	[8]
<i>Drosophila guanche</i>	SGM-IS	SGM Transposon Family	Class II	non-autonomous	-	SGM satDNA	[85]
<i>Drosophila virilis</i> group	pDv element	pDv transposable element family	Class II	-	Terminal repeat	pvB370 BamHI sat DNA	[86]
<i>Drosophila melanogaster</i>	TART   HeT-A	TART subfamilies of the HeT DNA family	Class I Class I	Autonomous non-autonomous	-3' UTR	18HT satDNA	[87]
<i>Drosophila virilis</i>	Tetris	Foldback elements	Class II	non-autonomous	TIR	satDNA-arrays (TIR-220)	[88]
<i>Hydromantes imperialis</i> and <i>H. ambrosii</i>	SINE-like elements	SINE Superfamily	Class I	non-autonomous	tRNA-related region	Hy/Pol III	[89]
<i>Monopterus albus</i>	LTR-Retrotransposon Gypsy	LTR-Retrotransposon-like	Class I	autonomous	LTR	satDNA MALREP	[90]
<i>Xenopus leavis</i>	Xmix MITE	-	Class II	non-autonomous	TIR	Xstir satDNA	[73]
<i>Xenopus leavis</i>	SINE-like	SINE Superfamily	Class I	non-autonomous	tRNA-related region	Satellite 1	[91]
<i>Xenopus tropicalis</i>	MITE of TC1-mariner	Tc1-Mariner	Class II	non-autonomous	stDNA located within the MITE element	miDNA4	[74]
<i>Ostrea edulis</i>	CvA	Pearl	Class II	non-autonomous	-	HindIII	[77]
<i>Venerupis decussata</i>	MITE (Pearl)	Pearl	Class II	non-autonomous	-	BIV160	[92]
<i>Arabidopsis thaliana</i>	Atenspm	En/Spm-like	Class II	autonomous	-5' UTR	ENSAT1	[93]
<i>Pisum sativum</i>	Ty3/gypsy-like ogre	Gypsy	Class I	autonomous	-3' UTR	PisTR-A satDNA	[59]
<i>Glycine max</i>	Gmr9/GmOgre	Ty3-gypsy	Class I	autonomous	Between the 3'UTR and Repetitive LTR	Gmr9-associate minisatellite	[94]

Table 1. Cont.

Species	Type of Transposable Element (TE)	Superfamily	Class	Mobility	Region of the TE	New Sequence	Reference
<i>Solanum bulbocastanum</i>	Sore 1	SORE-1 family	Class I	autonomous	LTR	Sobo satDNA	[6]
<i>Solanum tuberosum</i>	Ty3/gypsy-like	LTR-Retrotransposon-like	Class I	autonomous	LTR	St3-58; St3-238; St18; St3-294	[95]
<i>Zea mays</i>	CRM1 and CRM4	Centromeric Retrotransposons of Maize (CRM)	Class I	autonomous	UTR regions and LTR	CRM1TR e CRM4TR satDNA	[96]
<i>Hordeum vulgare</i>	BARE-1	BARE-1 retrotransposon family	Class I	non-autonomous	LTR	SSR	[97]
<i>Aegilops speltoides</i>	Cereba	Ty3-gypsy	Class I	autonomous	ORF to gag	CAA microsatellite	[98]
<i>Secale cereale</i>	Crwydryn Tnr1 MITE	Crwydryn Tnr1/Stowaway family	Class I Class II	non-autonomous non-autonomous	ORF to gag -	E3900 satDNA D1100 satDNA	[75]
<i>Lathyrus sativus</i>	Ogre LTR-retrotransposons	LTR-Retrotransposon-like	Class I	autonomous	Close to the ORF to gag	nine satDNAs.	[99]
<i>Chenopodium album</i> aggregate	CACTA-Like Jozin	CACTA superfamily	Class II	non-autonomous	Tnp2 TPase	CficCl-61-40 satDNA	[4]

The MITE sequences experience burst events, which may lead to a dramatic increase in the number of copies, resulting in the rapid accumulation of arranged units with highly similar sequences in the chromosomes [100–102], as seen in the amphibian *Xenopus tropicalis* [74]. In the genome of *X. tropicalis*, the presence of a miDNA4-MITE that contains a satellite DNA motif, demonstrates the synergic interplay between the MITE structure and genomic context, which contributes to the emergence of the repetitive arrays of the new satDNA. In this model, the birth of the satellite monomer within a MITE is followed by the amplification of tandem repeats, with the higher recombination rates of the tandem arrays of paralogous copies leading to the rapid homogenization of the repetitive arrays and, over time, the generation of satDNA following the putative concerted evolution model [74].

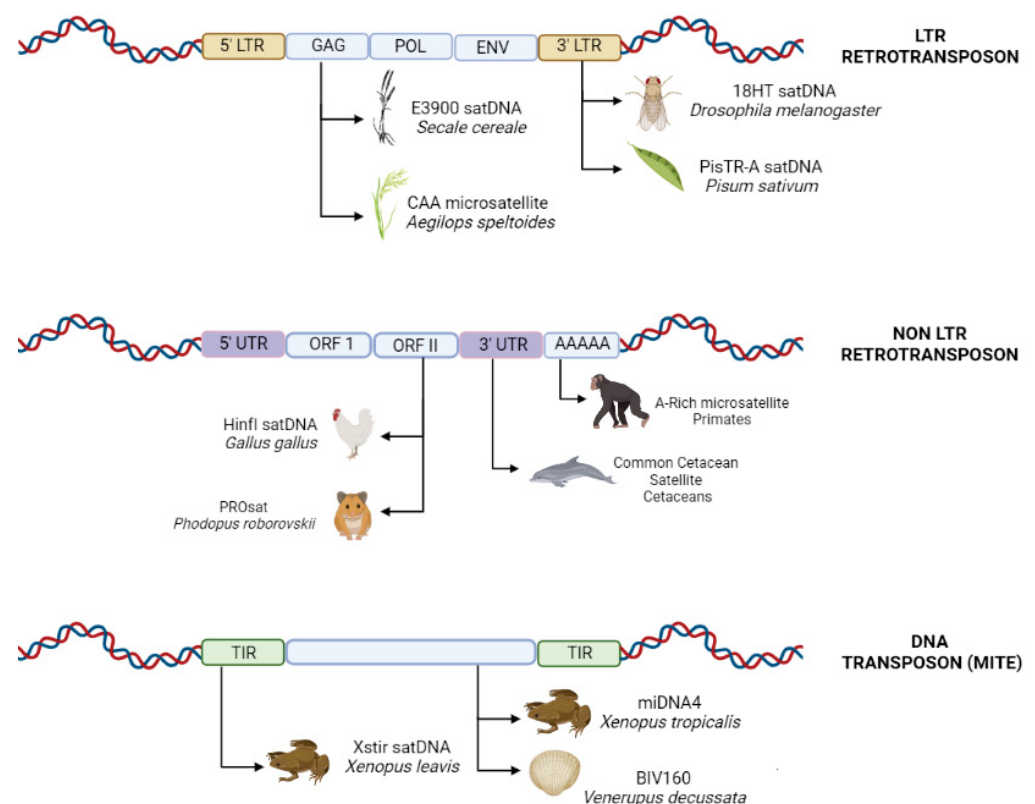
Helitron elements (subclass II of the DNA transposons) also represent a source of the spread of satDNA-like arrays in the genome. The transposition of these Class II elements occurs through semi-replicative transposition, in which only one strand of the transposon is transferred between genomic sites without duplicating the target site (TSD) upon insertion, a process known as Rolling Circle Transposition, or RCT [2,9,103,104]. Analysis of the RCT mechanism has revealed cases in which the rolling-circle transposases are unable to recognize the termination sites located at the extremity of each TE, which results in the transposition of fragments of the genomic DNA located in the immediate vicinity of the sequence. This makes the TEs prone to the capture and propagation of a range of different genomic sequences [2,105].

As satDNA may be formed by the tandem amplification of a whole TE, or only a part of it, where fractions of short satDNA-like arrays would be expected to be found dispersed throughout the genome as an intermediate stage of the emergence of satDNA from the TEs, which are normally distributed in euchromatic regions [2,8,74,106]. In *Crassostrea gigas*,

for example, the genome assembly presents 13 clusters of satDNA-like tandem repeats, which represent the central repeats of 11 non-autonomous elements belonging to the Helentron superfamily of DNA transposons known as the CgHINE [2]. The genome-wide distribution of this element in this species indicates that Helentrons are able to propagate tandem repeats.

## 2.2. Are Certain Portions of a TE More Prone to the Generation of Tandem Repeat Sequences?

Although the entire sequence of a transposable element has the potential to act as the substrate for the generation of new repetitive elements [61], we observed a prevalence of the repetitive portions of these elements as the source of the monomeric units of tandem repeats. The greater similarities between the monomeric units and the Terminal Inverted Repeats (TIRs), Long Terminal Repeats (LTRs), and other non-coding regions of the transposable elements highlights the importance of these naturally repetitive segments for the emergence of new classes of repetitive DNA (Table 1; Figure 2).



**Figure 2.** Schematic diagram of the different regions of transposable elements that may provide the starting point of origin for new micro-, mini- or satellite DNA. **LTR** (Long Terminal Repeat); **GAG** (GAG domain); **POL** (Reverse Transcriptase); **ENV** (Envelope Protein); **UTR** (Untranslated Region); **ORF** (Open Reading Frame); **TIR** (Terminal Inverted Repeats). (Source: the authors/created in [Biorender.com](https://www.biorender.com), accessed on 5 September 2022).

The untranslated 5' and 3' (5'UTR and 3'UTR) extremities also serve as a source of satellite DNA. The evidence available for the pea, *Pisum sativum*, indicates that the satDNA PisTR-A originated through the amplification and homogenization of tandem repeats present in the hypervariable 3'UTR of the Ty3/gypsy-like Ogre elements [59]. The 3' terminal may also play a key role in this process, as observed in *Drosophila melanogaster*, in which the emergence of new satDNA corresponded to the 3' non-coding region of the transposable element HeT-A [87], and in the cetaceans, the DNA of the common satellite monomer is similar to the 3'-terminal portion of the mammalian L1 (LINE-1) retrotransposon [82] (Figure 2).

Open Reading Frames (ORFs) of TEs have also been found to be related to the satDNA in *Gallus gallus*, with similarities in the sequences being observed among the regions of the satDNA *Hinf*I and the CR1 retrotransposon, which contains the partial ORF II of the CR1 element [7] (Figure 2). In the rodent *Phodopus roborovskii*, the ORF II of a LINE-1 retrotransposon is 88% similar to a small region of the PROsat present in the genome [81]. In rye, *Secale cereale*, the E900 satDNA contains a fragment of a retrotransposon which encodes a partial reading frame for the GAG-like protein of an LTR retrotransposon [75]. The GAG-like gene was also related to the CAA microsatellite in wheat, *Aegilops speltoides*, and presented similarities with the upstream region of the Ty3/gypsy-like retroelement [98] (Figure 2). Overall, then, the sum of the evidence indicates that any part of a transposable element may provide a substrate for the generation of new sequences, and there does not appear to be any conclusive evidence that specific portions of these elements are involved preferentially in the generation of satellites or other types of tandem repeat.

### 3. Is the Centromeric Region a Hotspot of the Emergence of de novo satDNA Derived from TEs?

Centromeric regions are favorable to the emergence and establishment of new families of satellite DNA and are ideal models for studying the TEs as an evolutionary substrate in chromosomal evolution [107,108]. Centromeres are gene-poor, and the majority of the transcriptional activity observed in this region involves non-coding RNAs that interact in the organization of the kinetochores [109].

Meiotic recombination is avoided in centromeric regions, which appears to be an evolutionary strategy to avoid chromosome aneuploidy, given that crossovers in the vicinity of the centromere make chromosomes more prone to mis-segregation [109,110]. The centromeres are thus a cold spot for crossovers, making this chromosomal region prone to the emergence of de novo satDNA from TEs. The mobile genetic elements inserted in these regions cannot be deleted easily by crossovers, and thus accumulate in the centromeric and pericentromeric regions [111].

Plant and animal centromeres are rich in TEs, and these sequences may sometimes be highly specific and/or involved directly in the architecture of this chromosomal region. Two species of beetle, *Dichotomius schiffleri* and *Tribolium castaneum*, have unusually extended centromeres, which appear to be related to the prevalence of DsGypsy6, LINE-1, and Helitron-like sequences in these structures [112,113]. In Poaceae species, the centromere is enriched with a specific retrotransposon family known as the centromeric retrotransposon (CR), which has highly conserved motifs, that are also found in their B chromosomes, when present (see [114]). The CR elements and centromere connection may contribute to the maintenance of the centromere/kinetochore complex, given that these sequences interact with the kinetochore protein CENH3 [115].

Recent studies of neo-centromeres have established a new perspective for the understanding of the role of TEs in the emergence of new satDNA families. Talbot and Hennikof [111] proposed a model which indicates an active role of the TEs in the evolutionary emergence of neo-centromeres. De novo centromeres appear to arise from the deposition of epigenetic markers in a particular region of the chromosome, followed by the enrichment of repetitive sequences at this new site [116]. The transposons appear to play an important role in the maintenance of the size of the region and the increase in the repeat content of the neo-centromeres that do not have many tandem repeats during this initial period [117].

Large numbers of CENP-A are detected flanking centromeres, but low levels of CENP-A may spread to the non-canonical centromere regions, predisposing them to acquire a potential centromeric function [118,119], and the insertion of TEs in an ectopic CENP-A site may create favorable conditions for the evolution of new a centromeric complex [118].

New cycles of transposition activity increase the density of TEs and may result in insertions of copies adjacent to the original, resulting in tandem duplications (tandem dimers, trimers, or more repetitions), as observed in the Rice8 satDNA in centromere of rice [120],

for example. Retrotransposons comprise approximately 70% of the functional centromeres of *D. melanogaster*, and are composed of complex DNA rich in non-LTR retroelements inserted within large blocks of tandem repeats [121]. Some plant centromeres are composed of long arrays of satDNA interspersed with Gypsy LTR-retrotransposons [122–125], which highlights the interplay of the different classes of repetitive DNA in the establishment and maintenance of the centromere region.

Satellite DNA repeats derived from TEs in the centromere region have been identified in a number of different species [3,61,121,126]. The centromeres of *Zea mays* are a good example because it is composed of CRM1TR and CRM4TR satDNA, two tandem arrays that were derived from an LTR-retrotransposon in at least two separate events [96]. The repeat arrays of satellites in the centromeres of chromosomes 1, 2, 3, 5, 7, and 8 of the potato, *Solanum tuberosus*, were also amplified from retrotransposon sequences [95]. In the chicken karyotype, several centromere-specific types of satDNA are highly similar to retrotransposons [127]. In the centromeres of *Prunus* species, a highly conserved monomer unit of 166 bps has been identified from assembled genomes and sequencing reads, with varying signal intensities in fluorescence in situ hybridization (FISH) experiments, which indicates that the centromeric regions of this genus are enriched with this sequence [128].

The centromeric region of the Y chromosome of *Drosophila melanogaster* is composed of 18HT satDNA, a satellite DNA derived from HeT-A and TART non-LTR retrotransposons [87], the same elements that were co-opted to the telomeric function in *Drosophila*. In the *Drosophila obscura* species group (*Drosophila subobscura*, *D. guanche*, and *D. madeirensis*), a DNA transposon (SGM elements) was the substrate for the emergence of a species-specific SGM satDNA that makes up the centromeric heterochromatin in *D. guanche*, in which the SGM element appears to be inactive [85].

The subtelomeric heterochromatin also appears to be prone to the emergence of tandem copies of transposable elements. *Drosophila melanogaster*, for example, has three telomere-specialized retrotransposons (HeT-A, TART, and TAHRE) that have been co-opted functionally to the organization of the chromatin and the maintenance of the telomere, and resemble telomere extensions containing telomerase [129,130]. The subtelomeric region of the *D. melanogaster* chromosomes presents a complex combination of potentially active elements and truncated TEs arranged in a long array [129]. In *Drosophila biarmipes*, by contrast, Helitron transposons may play an important role in the structure of the telomere, and while full-length Helitrons can be observed in the telomeres, fragments are interspersed within the abundant satellite sequences [129]. The richness of the repetitive arrays derived from the TEs in the subtelomeric heterochromatin of *Drosophila* represents a promising model for the study of the emergence of satellite DNA from TE elements and another potential chromosomal hotspot for the formation of de novo satDNA from TEs.

#### 4. Mechanisms of the Production of Repeats from TEs

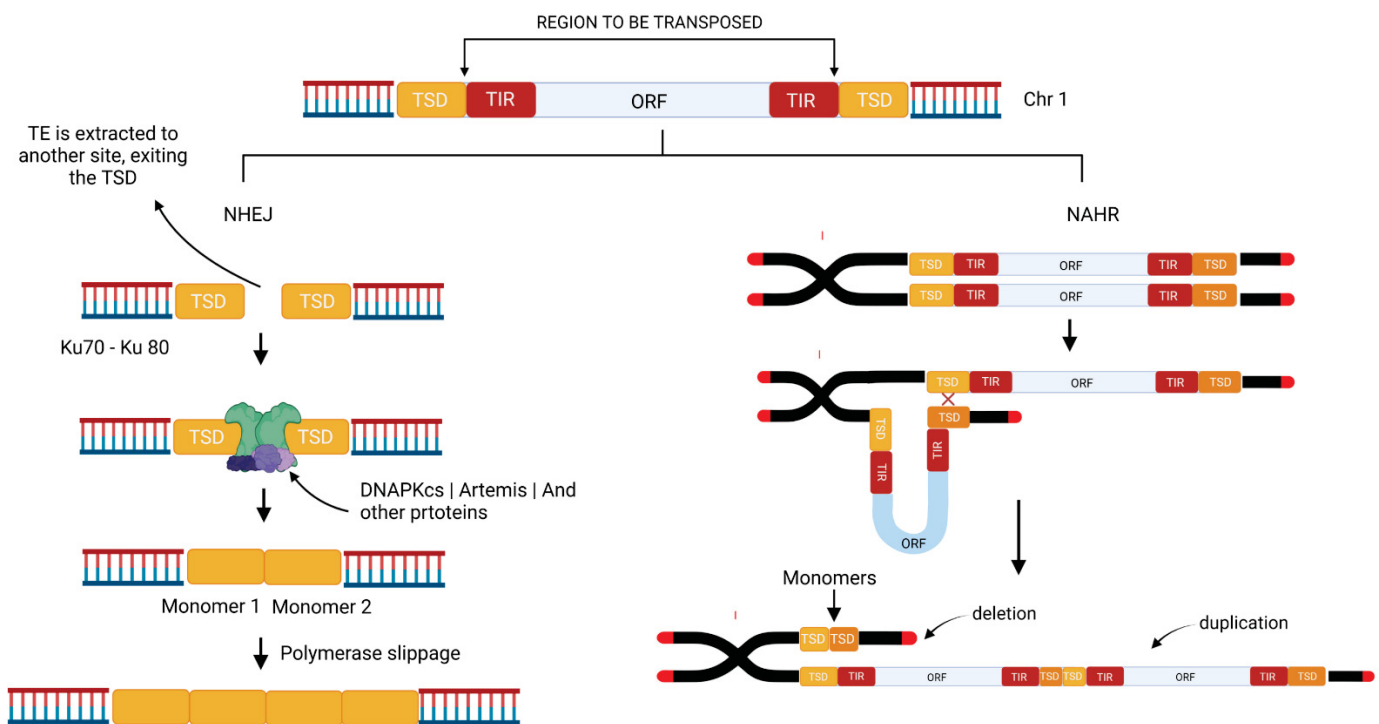
Unequal crossovers are an important mechanism of the expansion of tandem arrays and the homogenization of the satDNA by concerted evolution [131], although the absence of canonical meiotic crossovers in the centromeric region would require a mechanism that is independent of the emergence and evolution of the satDNA in this region. The transposition process may cause a Double Strand Break (DSB) of the DNA, which the DNA repair mechanism will attempt to fix, through alternative pathways, such as Non-Homologous End Joining (NHEJ), Non-Allelic Homologous Recombination (NAHR), or by Homologous Recombination (HR).

The NHEJ may contribute to the expansion of tandem arrays as a result of a single transposition event. Microhomology-mediated NHEJ has the potential to give rise to a wide range of chromosome events including duplications of sequences if it happens between sister chromatids [132]. This mechanism also plays an important role in finalizing some TE-related instability events [133] and has been found with a dominant role in gene duplications [134]. TSD sequences are identical; therefore, it provides microhomology



making it possible for NHEJ to restore the genome back to its original state before the insertion of a transposable element [135].

For example, Target Site Duplications (TSD) may be brought together by the NHEJ mechanism following a “cut and paste” transposition event (Figure 3), which rejoins the broken ends without the use of extensive homology [136,137]. As each TSD normally has fewer than 10 bps, these dimers of low complexity can arise easily through polymerase slippage and the subsequent expansion of the number of copies (Figure 3).



**Figure 3.** Schematic diagram showing how the Non-Homologous End-Joining (NHEJ) and Non-Allelic Homologous Recombination (NAHR) DNA repair mechanisms contribute to the expansion of tandem repeats from a Transposable Element. **TSD** (Target Site Duplication); **TIR** (Terminal Inverted Repeats); **ORF** (Open Reading Frame). (Source: Authors/created in [Biorender.com](https://www.biorender.com), accessed on 5 September 2022).

Studies on TSD and TE insertions were performed in *D. melanogaster* [1], where it was hypothesized that TEs organized in tandem could be generated through multiple insertions at the same chromosomal site, which also presents a TSD. The authors concluded, analyzing reads of the element known as P-element, that the junctions of the tandem repeats had the same size as the TSD of the said element, and the consensus motif was found similar to the TSD from other P-element insertions.

On the other hand, NAHR may also result in recurrent genome rearrangements, such as inversions, translocations, deletions, or duplications, in sequences that are not in intra- or inter-chromatid allelic positions, i.e., they are paralogous [138,139]. This mechanism provides valuable insights into the expansion of repeat arrays to the emergence of new satDNA, given that it depends on the position of the paralogous TE copies that will be used as the template for the repair of the chromatid that has suffered the DSB, which may result in tandem duplications of the repetitive segments of the TE (Figure 3). Given this, intra- or inter-chromatid NAHR may contribute to the expansion of the initial repeat arrays in a manner similar to that of the unequal crossovers performed in the concerted evolution model.

One other especially interesting pathway of double-strand DNA repair is the use of a strand from a sister chromatid through Homologous Recombination (HR). Given either the

existence of the repetitive elements that make up the molecular structure of the TEs (such as LTRs and ITRs) or the greater similarities of the TSD regions, the repair of DBSs using intrachromosomal HR mechanisms has the potential to generate tandem duplications in these regions [140] which may cause repeat arrays of the TE to accumulate in this region, if the Holliday junctions are processed to yield a crossover.

Similar results may also be obtained when a DBS is repaired by ectopic recombination, in which the paralogous copies used as the template are from non-homologous chromosomes [141]. In this case, the resolution of the Holliday junctions in the crossover events between dispersed repeats [142] will also result in the tandem duplications of the repetitive portions of the TEs in the region of the DBS. The outcome of these mechanisms of DBS repair is the formation of repeat arrays in the region of the DNA damage, which will be a source of other mechanisms of sequence expansion, such as polymerase slippage, that contribute to an increase in the number of monomers, leading to the emergence of satDNA from TE segments.

Finally, the intrinsic mechanisms involved in the transposition events are prone to the production of repeats derived from TEs. For example, the genesis of the Bari1 (Tc1-mariner superfamily) repeat clusters in *D. melanogaster* was provoked by anomalous rolling circle mechanisms and the subsequent reintegration within the *Stalker* LTR-retrotransposon, which generated an 80-repeat organized in heterochromatic clusters [143].

## 5. Conclusions

Repetitive DNA sequences have long been studied because of their structural role and impact on the genome of a range of different organisms. Here, however, we highlight the importance of also focusing on the relationships among the different elements. The possible presence of pre-existing transposable elements that serve as a substrate for the emergence of new families of repetitive sequences reinforces the potential importance of these elements in genome and karyotype remodeling during the evolutionary process. Given this, a better comprehension of the evolutionary relationships of these elements may be extremely valuable for a better understanding of the evolution of the eukaryote genome.

**Author Contributions:** M.L.Z. drafted and revised the manuscript. D.P.B. coordinated the study and revised the text. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Brazilian Coordination for Higher Education Personnel Training (CAPES/PROAP—Finance Code 001) and the Brazilian National Council for Scientific and Technological Development (CNPq #303646/2021-7; CNPq/MCTI/FNDCT N° 18/2021 #405041/2021-7).

**Acknowledgments:** We thank the Brazilian Coordination for Higher Education Personnel Training (CAPES/PROAP—Finance Code 001) for the scholarships provided to MLZ. DPB thanks the Brazilian National Council for Scientific and Technological Development (CNPq #303646/2021-7; CNPq/MCTI/FNDCT N° 18/2021 #405041/2021-7), for its financial support.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

1. McGurk, M.P.; Barbash, D.A. Double Insertion of Transposable Elements Provides a Substrate for the Evolution of Satellite DNA. *Genome Res.* **2018**, *28*, 714–725. [\[CrossRef\]](#)
2. Zeljko, V.T.; Pavlek, M.; Meštrović, N.; Plohl, M. Satellite DNA-like Repeats Are Dispersed throughout the Genome of the Pacific Oyster *Crassostrea Gigas* Carried by Helentron Non-Autonomous Mobile Elements. *Sci. Rep.* **2020**, *10*, 15107. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Meštrović, N.; Mravinac, B.; Pavlek, M.; Vojvoda-Zeljko, T.; Šatović, E.; Plohl, M. Structural and Functional Liaisons between Transposable Elements and Satellite DNAs. *Chromosome Res.* **2015**, *23*, 583–596. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Belyayev, A.; Josefiová, J.; Jandová, M.; Mahelka, V.; Krak, K.; Mandák, B. Transposons and Satellite DNA: On the Origin of the Major Satellite DNA Family in the *Chenopodium* Genome. *Mob. DNA* **2020**, *11*, 20. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Paço, A.; Freitas, R.; Vieira-da-Silva, A. Conversion of DNA Sequences: From a Transposable Element to a Tandem Repeat or to a Gene. *Genes* **2019**, *10*, 1014. [\[CrossRef\]](#)
6. Tek, A.L.; Song, J.; Macas, J.; Jiang, J. Sobo, a Recently Amplified Satellite Repeat of Potato, and Its Implications for the Origin of Tandemly Repeated Sequences. *Genetics* **2005**, *170*, 1231–1238. [\[CrossRef\]](#)

7. Li, J.; Leung, F.C. A CR1 Element Is Embedded in a Novel Tandem Repeat (*Hin* FI Repeat) within the Chicken Genome. *Genome* **2006**, *49*, 97–103. [[CrossRef](#)]
8. Dias, G.B.; Heringer, P.; Svartman, M.; Kuhn, G.C.S. Helitrons Shaping the Genomic Architecture of *Drosophila*: Enrichment of DINE-TR1 in  $\alpha$ - and  $\beta$ -Heterochromatin, Satellite DNA Emergence, and PiRNA Expression. *Chromosome Res.* **2015**, *23*, 597–613. [[CrossRef](#)]
9. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A Unified Classification System for Eukaryotic Transposable Elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)]
10. Kidwell, M.G.; Lisch, D.R. Perspective: Transposable elements, parasitic dna, and genome evolution. *Evolution* **2001**, *55*, 1–24. [[CrossRef](#)]
11. Fernández-Medina, R.D.; Ribeiro, J.M.C.; Carareto, C.M.A.; Velasque, L.; Struchiner, C.J. Losing Identity: Structural Diversity of Transposable Elements Belonging to Different Classes in the Genome of *Anopheles Gambiae*. *BMC Genom.* **2012**, *13*, 272. [[CrossRef](#)] [[PubMed](#)]
12. Hartl, D.L.; Lozovskaya, E.R.; Nurminsky, D.I.; Lohe, A.R. What Restricts the Activity of Mariner-like Transposable Elements. *Trends Genet.* **1997**, *13*, 197–201. [[CrossRef](#)]
13. Le Rouzic, A.; Capy, P. The First Steps of Transposable Elements Invasion. *Genetics* **2005**, *169*, 1033–1043. [[CrossRef](#)] [[PubMed](#)]
14. Oliveira, S.G.; Bao, W.; Martins, C.; Jurka, J. Horizontal Transfers of Mariner Transposons between Mammals and Insects. *Mob. DNA* **2012**, *3*, 14. [[CrossRef](#)] [[PubMed](#)]
15. Pace, J.K.; Feschotte, C. The Evolutionary History of Human DNA Transposons: Evidence for Intense Activity in the Primate Lineage. *Genome Res.* **2007**, *17*, 422–432. [[CrossRef](#)]
16. Pritham, E.J.; Feschotte, C. Massive Amplification of Rolling-Circle Transposons in the Lineage of the Bat *Myotis Lucifugus*. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1895–1900. [[CrossRef](#)]
17. Chen, J.; Lu, L.; Robb, S.M.C.; Collin, M.; Okumoto, Y.; Stajich, J.E.; Wessler, S.R. Genomic Diversity Generated by a Transposable Element Burst in a Rice Recombinant Inbred Population. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 26288–26297. [[CrossRef](#)]
18. Nitta, N.; Farman, M.L.; Leong, S.A. Genome Organization of *Magnaporthe Grisea*: Integration of Genetic Maps, Clustering of Transposable Elements and Identification of Genome Duplications and Rearrangements. *Theor. Appl. Genet.* **1997**, *95*, 20–32. [[CrossRef](#)]
19. Dasilva, C.; Hadji, H.; Ozouf-Costaz, C.; Nicaud, S.; Jaillon, O.; Weissenbach, J.; Crollius, H.R. Remarkable Compartmentalization of Transposable Elements and Pseudogenes in the Heterochromatin of the *Tetraodon Nigroviridis* Genome. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13636–13641. [[CrossRef](#)]
20. Thon, M.R.; Martin, S.L.; Goff, S.; Wing, R.A.; Dean, R.A. BAC End Sequences and a Physical Map Reveal Transposable Element Content and Clustering Patterns in the Genome of *Magnaporthe Grisea*. *Fungal Genet. Biol.* **2004**, *41*, 657–666. [[CrossRef](#)]
21. Thon, M.; Pan, H.; Diener, S.; Papalás, J.; Taro, A.; Mitchell, T.; Dean, R. BAC End Sequences and a Physical Map Reveal Transposable Element Content and Clustering Patterns in the Genome of *Magnaporthe Grisea*. *Genome Biol.* **2006**, *7*, R16. [[CrossRef](#)] [[PubMed](#)]
22. Zamudio, N.; Bourc'his, D. Transposable Elements in the Mammalian Germline: A Comfortable Niche or a Deadly Trap? *Heredity* **2010**, *105*, 92–104. [[CrossRef](#)] [[PubMed](#)]
23. Bao, J.; Yan, W. Male Germline Control of Transposable Elements. *Biol. Reprod.* **2012**, *86*, 162–1. [[CrossRef](#)]
24. Shi, X.; Seluanov, A.; Gorbunova, V. Cell Divisions Are Required for L1 Retrotransposition. *Mol. Cell. Biol.* **2007**, *27*, 1264–1270. [[CrossRef](#)] [[PubMed](#)]
25. Ernst, C.; Odom, D.T.; Kutter, C. The Emergence of PiRNAs against Transposon Invasion to Preserve Mammalian Genome Integrity. *Nat. Commun.* **2017**, *8*, 1411. [[CrossRef](#)]
26. Feschotte, C.; Pritham, E.J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* **2007**, *41*, 331–368. [[CrossRef](#)]
27. Engels, W.R.; Johnson-Schlitz, D.M.; Eggleston, W.B.; Sved, J. High-Frequency P Element Loss in *Drosophila* Is Homolog Dependent. *Cell* **1990**, *62*, 515–525. [[CrossRef](#)]
28. Maupetit-Mehouas, S.; Vaury, C. Transposon Reactivation in the Germline May Be Useful for Both Transposons and Their Host Genomes. *Cells* **2020**, *9*, 1172. [[CrossRef](#)]
29. Sigman, M.J.; Slotkin, R.K. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *Plant Cell* **2016**, *28*, 304–313. [[CrossRef](#)]
30. Ward, M.C.; Zhao, S.; Luo, K.; Pavlovic, B.J.; Karimi, M.M.; Stephens, M.; Gilad, Y. Silencing of Transposable Elements May Not Be a Major Driver of Regulatory Evolution in Primate iPSCs. *eLife* **2018**, *7*, e33084. [[CrossRef](#)]
31. Rigal, M.; Mathieu, O. A “Mille-Feuille” of Silencing: Epigenetic Control of Transposable Elements. *Biochim. Biophys. Acta (BBA)—Gene Regul. Mech.* **2011**, *1809*, 452–458. [[CrossRef](#)] [[PubMed](#)]
32. Yang, F.; Xi, R. Silencing Transposable Elements in the *Drosophila* Germline. *Cell. Mol. Life Sci.* **2017**, *74*, 435–448. [[CrossRef](#)] [[PubMed](#)]
33. Moore, L.D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38*, 23–38. [[CrossRef](#)] [[PubMed](#)]
34. Chandler, V.; Rivin, C.; Walbot, V. Stable non-mutator stocks of maize have sequences homologous to the *Mu1* transposable element. *Genetics* **1986**, *114*, 1007–1021. [[CrossRef](#)]

35. Chomet, P.S.; Wessler, S.; Dellaporta, S.L. Inactivation of the Maize Transposable Element Activator (Ac) Is Associated with Its DNA Modification. *EMBO J.* **1987**, *6*, 295–302. [[CrossRef](#)]
36. Banks, J.A.; Masson, P.; Fedoroff, N. Molecular Mechanisms in the Developmental Regulation of the Maize Suppressor-Mutator Transposable Element. *Genes Dev.* **1988**, *2*, 1364–1380. [[CrossRef](#)]
37. Collings, C.K.; Anderson, J.N. Links between DNA Methylation and Nucleosome Occupancy in the Human Genome. *Epigenetics Chromatin* **2017**, *10*, 18. [[CrossRef](#)]
38. Rose, N.R.; Klose, R.J. Understanding the Relationship between DNA Methylation and Histone Lysine Methylation. *Biochim. Biophys. Acta Gene Regul. Mech.* **2014**, *1839*, 1362–1372. [[CrossRef](#)]
39. Walter, M.; Teissandier, A.; Pérez-Palacios, R.; Bourc'his, D. An Epigenetic Switch Ensures Transposon Repression upon Dynamic Loss of DNA Methylation in Embryonic Stem Cells. *eLife* **2016**, *5*, e11418. [[CrossRef](#)]
40. Ebbs, M.L.; Bender, J. Locus-Specific Control of DNA Methylation by the Arabidopsis SUVH5 Histone Methyltransferase. *Plant Cell* **2006**, *18*, 1166–1176. [[CrossRef](#)]
41. Lippman, Z.; May, B.; Yordan, C.; Singer, T.; Martienssen, R. Distinct Mechanisms Determine Transposon Inheritance and Methylation via Small Interfering RNA and Histone Modification. *PLoS Biol.* **2003**, *1*, E67. [[CrossRef](#)] [[PubMed](#)]
42. Janssen, A.; Colmenares, S.U.; Karpen, G.H. Heterochromatin: Guardian of the Genome. *Annu. Rev. Cell Dev. Biol.* **2018**, *34*, 265–288. [[CrossRef](#)] [[PubMed](#)]
43. Liu, J.; Ali, M.; Zhou, Q. Establishment and Evolution of Heterochromatin. *Ann. N. Y. Acad. Sci.* **2020**, *1476*, 59–77. [[CrossRef](#)] [[PubMed](#)]
44. Kidwell, M.G. Transposable Elements and the Evolution of Genome Size in Eukaryotes. *Genetica* **2002**, *115*, 49–63. [[CrossRef](#)]
45. Sirijovski, N.; Woolnough, C.; Rock, J.; Joss, J.M.P. NfCR1, the First Non-LTR Retrotransposon Characterized in the Australian Lungfish Genome, *Neoceratodus Forsteri*, Shows Similarities to CR1-like Elements. *J. Exp. Zool.* **2005**, *304*, 40–49. [[CrossRef](#)] [[PubMed](#)]
46. Wang, K.; Wang, J.; Zhu, C.; Yang, L.; Ren, Y.; Ruan, J.; Fan, G.; Hu, J.; Xu, W.; Bi, X.; et al. African Lungfish Genome Sheds Light on the Vertebrate Water-to-Land Transition. *Cell* **2021**, *184*, 1362–1376.e18. [[CrossRef](#)] [[PubMed](#)]
47. Marsano, R.M.; Dimitri, P. Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements. *Cells* **2022**, *11*, 761. [[CrossRef](#)]
48. Kapitonov, V.V.; Jurka, J. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biol.* **2005**, *3*, e181. [[CrossRef](#)]
49. Feschotte, C. Transposable Elements and the Evolution of Regulatory Networks. *Nat. Rev. Genet.* **2008**, *9*, 397–405. [[CrossRef](#)]
50. Alzohairy, A.M.; Gyulai, G.; Jansen, R.K.; Bahieldin, A. Transposable Elements Domesticated and Neofunctionalized by Eukaryotic Genomes. *Plasmid* **2013**, *69*, 1–15. [[CrossRef](#)]
51. Qiu, Y.; Köhler, C. Mobility Connects: Transposable Elements Wire New Transcriptional Networks by Transferring Transcription Factor Binding Motifs. *Biochem. Soc. Trans.* **2020**, *48*, 1005–1017. [[CrossRef](#)] [[PubMed](#)]
52. Ali, A.; Han, K.; Liang, P. Role of Transposable Elements in Gene Regulation in the Human Genome. *Life* **2021**, *11*, 118. [[CrossRef](#)] [[PubMed](#)]
53. Fueyo, R.; Judd, J.; Feschotte, C.; Wysocka, J. Roles of Transposable Elements in the Regulation of Mammalian Transcription. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 481–497. [[CrossRef](#)] [[PubMed](#)]
54. Kelkar, Y.D.; Eckert, K.A.; Chiaromonte, F.; Makova, K.D. A Matter of Life or Death: How Microsatellites Emerge in and Vanish from the Human Genome. *Genome Res.* **2011**, *21*, 2038–2048. [[CrossRef](#)]
55. Etchegaray, E.; Naville, M.; Volff, J.-N.; Haftek-Terreau, Z. Transposable Element-Derived Sequences in Vertebrate Development. *Mob. DNA* **2021**, *12*, 1. [[CrossRef](#)]
56. Slotkin, R.K.; Martienssen, R. Transposable Elements and the Epigenetic Regulation of the Genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285. [[CrossRef](#)]
57. Elbarbary, R.A.; Lucas, B.A.; Maquat, L.E. Retrotransposons as Regulators of Gene Expression. *Science* **2016**, *351*, aac7247. [[CrossRef](#)]
58. Buschiazzo, E.; Gemmell, N.J. The Rise, Fall and Renaissance of Microsatellites in Eukaryotic Genomes—PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/16998838/> (accessed on 15 May 2022).
59. Macas, J.; Koblízková, A.; Navrátilová, A.; Neumann, P. Hypervariable 3' UTR Region of Plant LTR-Retrotransposons as a Source of Novel Satellite Repeats. *Gene* **2009**, *448*, 198–206. [[CrossRef](#)]
60. Kazazian, H.H. Mobile Elements: Drivers of Genome Evolution. *Science* **2004**, *303*, 1626–1632. [[CrossRef](#)]
61. Ahmed, M.; Liang, P. Transposable Elements Are a Significant Contributor to Tandem Repeats in the Human Genome. *Comp. Funct. Genom.* **2012**, *2012*, 947089. [[CrossRef](#)]
62. Kanhayuwa, L.; Coutts, R.H.A. Short Interspersed Nuclear Element (SINE) Sequences in the Genome of the Human Pathogenic Fungus *Aspergillus Fumigatus* Af293. *PLoS ONE* **2016**, *11*, e0163215. [[CrossRef](#)]
63. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. International Human Genome Sequencing Consortium. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
64. Dewannieux, M.; Esnault, C.; Heidmann, T. LINE-Mediated Retrotransposition of Marked Alu Sequences. *Nat. Genet.* **2003**, *35*, 41–48. [[CrossRef](#)] [[PubMed](#)]

65. Deininger, P. Alu Elements: Know the SINEs. *Genome Biol.* **2011**, *12*, 236. [[CrossRef](#)]
66. Armour, J.A.L.; Wong, Z.; Wilson, V.; Royle, N.J.; Jeffreys, A.J. Sequences Flanking the Repeat Arrays of Human Minisatellites: Association with Tandem and Dispersed Repeat Elements. *Nucl. Acids Res.* **1989**, *17*, 4925–4936. [[CrossRef](#)] [[PubMed](#)]
67. Clark, R.M.; Dalglish, G.L.; Endres, D.; Gomez, M.; Taylor, J.; Bidichandani, S.I. Expansion of GAA Triplet Repeats in the Human Genome: Unique Origin of the FRDA Mutation at the Center of an Alu. *Genomics* **2004**, *83*, 373–383. [[CrossRef](#)] [[PubMed](#)]
68. Jurka, J.; Gentles, A.J. Origin and Diversification of Minisatellites Derived from Human Alu Sequences. *Gene* **2006**, *365*, 21–26. [[CrossRef](#)] [[PubMed](#)]
69. Arcot, S.S.; Wang, Z.; Weber, J.L.; Deininger, P.L.; Batzer, M.A. Alu Repeats: A Source for the Genesis of Primate Microsatellites. *Genomics* **1995**, *29*, 136–144. [[CrossRef](#)]
70. Bureau, T.E.; Wessler, S.R. Mobile Inverted-Repeat Elements of the Tourist Family Associated with the Genes of Many Cereal Grasses. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1411–1415. [[CrossRef](#)]
71. Feschotte, C.; Swamy, L.; Wessler, S.R. Genome-Wide Analysis of *Mariner*-Like Transposable Elements in Rice Reveals Complex Relationships with *Stowaway* Miniature Inverted Repeat Transposable Elements (MITEs). *Genetics* **2003**, *163*, 747–758. [[CrossRef](#)]
72. Coates, B.S.; Kroemer, J.A.; Sumerford, D.V.; Hellmich, R.L. A Novel Class of Miniature Inverted Repeat Transposable Elements (MITEs) That Contain Hitchhiking (GTCY)<sub>n</sub> Microsatellites: Lepidopteran Mobile Microsatellites. *Insect Mol. Biol.* **2011**, *20*, 15–27. [[CrossRef](#)] [[PubMed](#)]
73. Hikosaka, A.; Kawahara, A. Lineage-Specific Tandem Repeats Riding on a Transposable Element of MITE in *Xenopus* Evolution: A New Mechanism for Creating Simple Sequence Repeats. *J. Mol. Evol.* **2004**, *59*, 738–746. [[CrossRef](#)] [[PubMed](#)]
74. Scalvenzi, T.; Pollet, N. Insights on Genome Size Evolution from a Miniature Inverted Repeat Transposon Driving a Satellite DNA. *Mol. Phylogenet. Evol.* **2014**, *81*, 1–9. [[CrossRef](#)] [[PubMed](#)]
75. Langdon, T.; Seago, C.; Jones, R.N.; Ougham, H.; Thomas, H.; Forster, J.W.; Jenkins, G. De Novo Evolution of Satellite DNA on the Rye B Chromosome. *Genetics* **2000**, *154*, 869–884. [[CrossRef](#)] [[PubMed](#)]
76. López-Flores, I.; Garrido-Ramos, M.A. The Repetitive DNA Content of Eukaryotic Genomes. In *Genome Dynamics*; Garrido-Ramos, M.A., Ed.; S. Karger AG: Basel, Switzerland, 2012; Volume 7, pp. 1–28. [[CrossRef](#)]
77. López-Flores, I.; de la Herrán, R.; Garrido-Ramos, M.A.; Boudry, P.; Ruiz-Rejón, C.; Ruiz-Rejón, M. The Molecular Phylogeny of Oysters Based on a Satellite DNA Related to Transposons. *Gene* **2004**, *339*, 181–188. [[CrossRef](#)] [[PubMed](#)]
78. Bois, P.; Williamson, J.; Brown, J.; Dubrova, Y.E.; Jeffreys, A.J. A Novel Unstable Mouse VNTR Family Expanded from SINE B1 Elements. *Genomics* **1998**, *49*, 122–128. [[CrossRef](#)]
79. Kelly, R.G. Similar Origins of Two Mouse Minisatellites within Transposon-like LTRs. *Genomics* **1994**, *24*, 509–515. [[CrossRef](#)]
80. Rossi, M.S.; Pesce, C.G.; Reig, O.A.; Kornblihtt, A.R.; Zorzópulos, J. Retroviral-like Features in the Monomer of the Major Satellite DNA from the South American Rodents of the Genus *Ctenomys*. *DNA Sequence* **1993**, *3*, 379–381. [[CrossRef](#)]
81. Paço, A.; Adegas, F.; Meštrović, N.; Plohl, M.; Chaves, R. The Puzzling Character of Repetitive DNA in *Phodopus* Genomes (*Cricetidae*, *Rodentia*). *Chromosome Res.* **2015**, *23*, 427–440. [[CrossRef](#)]
82. Kapitonov, V.V.; Holmquist, G.P.; Jurka, J. L1 Repeat Is a Basic Unit of Heterochromatin Satellites in Cetaceans. *Mol. Biol. Evol.* **1998**, *15*, 611–612. [[CrossRef](#)]
83. Duffy, A.J.; Coltman, D.W.; Wright, J.M. Microsatellites at a Common Site in the Second ORF of L1 Elements in Mammalian Genomes. *Mamm. Genome* **1996**, *7*, 386–387. [[CrossRef](#)] [[PubMed](#)]
84. Palomeque, T.; Antonio Carrillo, J.; Muñoz-López, M.; Lorite, P. Detection of a *Mariner*-like Element and a Miniature Inverted-Repeat Transposable Element (MITE) Associated with the Heterochromatin from Ants of the Genus *Messor* and Their Possible Involvement for Satellite DNA Evolution. *Gene* **2006**, *371*, 194–205. [[CrossRef](#)] [[PubMed](#)]
85. Miller, W.J.; Nagel, A.; Bachmann, J.; Bachmann, L. Evolutionary Dynamics of the SGM Transposon Family in the *Drosophila Obscura* Species Group. *Mol. Biol. Evol.* **2000**, *17*, 1597–1609. [[CrossRef](#)]
86. Heikkinen, E.; Launonen, V.; Müller, E.; Bachmann, L. The PvB370 BamHI Satellite DNA Family of the *Drosophila Virilis* Group and Its Evolutionary Relation to Mobile Dispersed Genetic PDV Elements. *J. Mol. Evol.* **1995**, *41*, 604–614. [[CrossRef](#)] [[PubMed](#)]
87. Agudo, M.; Losada, A.; Abad, J.P.; Pimpinelli, S.; Ripoll, P.; Villasante, A. Centromeres from Telomeres? The Centromeric Region of the Y Chromosome of *Drosophila Melanogaster* Contains a Tandem Array of Telomeric HeT-A- and TART-Related Sequences. *Nucleic Acids Res.* **1999**, *27*, 3318–3324. [[CrossRef](#)] [[PubMed](#)]
88. Dias, G.B.; Svartman, M.; Delprat, A.; Ruiz, A.; Kuhn, G.C.S. Tetris Is a Foldback Transposon That Provided the Building Blocks for an Emerging Satellite DNA of *Drosophila Virilis*. *Genome Biol. Evol.* **2014**, *6*, 1302–1313. [[CrossRef](#)] [[PubMed](#)]
89. Batistoni, R.; Pesole, G.; Marracci, S.; Nardi, I. A Tandemly Repeated DNA Family Originated from SINE-Related Elements in the European Plethodontid Salamanders (*Amphibia*, *Urodela*). *J. Mol. Evol.* **1995**, *40*, 608–615. [[CrossRef](#)]
90. Suntronpong, A.; Singchat, W.; Kruasuwan, W.; Prakhongcheep, O.; Sillapaprayoon, S.; Muangmai, N.; Somyong, S.; Indananda, C.; Kraichak, E.; Peyachoknagul, S.; et al. Characterization of Centromeric Satellite DNAs (MALREP) in the Asian Swamp Eel (*Monopterus Albus*) Suggests the Possible Origin of Repeats from Transposable Elements. *Genomics* **2020**, *112*, 3097–3107. [[CrossRef](#)]
91. Pasero, P.; Sjakste, N.; Blettry, C.; Got, C.; Marilley, M. Long-Range Organization and Sequence-Directed Curvature of *Xenopus Laevis* Satellite 1 DNA. *Nucleic Acids Res.* **1993**, *21*, 4703–4710. [[CrossRef](#)]

92. Plohl, M.; Petrović, V.; Luchetti, A.; Ricci, A.; Šatović, E.; Passamonti, M.; Mantovani, B. Long-Term Conservation vs High Sequence Divergence: The Case of an Extraordinarily Old Satellite DNA in Bivalve Mollusks. *Heredity* **2010**, *104*, 543–551. [[CrossRef](#)]
93. Kapitonov, V.V.; Jurka, J. Molecular Paleontology of Transposable Elements from Arabidopsis Thaliana. In *Transposable Elements and Genome Evolution*; McDonald, J.F., Ed.; Springer: Dordrecht, The Netherlands, 2000; pp. 27–37. [[CrossRef](#)]
94. Mogil, L.S.; Slowikowski, K.; Laten, H.M. Computational and Experimental Analyses of Retrotransposon-Associated Minisatellite DNAs in the Soybean Genome. *BMC Bioinform.* **2012**, *13*, S13. [[CrossRef](#)] [[PubMed](#)]
95. Gong, Z.; Wu, Y.; Koblížková, A.; Torres, G.A.; Wang, K.; Iovene, M.; Neumann, P.; Zhang, W.; Novák, P.; Buell, C.R.; et al. Repeatless and Repeat-Based Centromeres in Potato: Implications for Centromere Evolution. *Plant Cell* **2012**, *24*, 3559–3574. [[CrossRef](#)]
96. Sharma, A.; Wolfgruber, T.K.; Presting, G.G. Tandem Repeats Derived from Centromeric Retrotransposons. *BMC Genom.* **2013**, *14*, 142. [[CrossRef](#)] [[PubMed](#)]
97. Ramsay, L.; Macaulay, M.; Cardle, L.; Morgante, M.; degli Ivanissevich, S.; Maestri, E.; Powell, W.; Waugh, R. Intimate Association of Microsatellite Repeats with Retrotransposons and Other Dispersed Repetitive Elements in Barley. *Plant J.* **1999**, *17*, 415–425. [[CrossRef](#)]
98. Cheng, Z.-J.; Murata, M. A Centromeric Tandem Repeat Family Originating From a Part of Ty3/ *Gypsy* -Retroelement in Wheat and Its Relatives. *Genetics* **2003**, *164*, 665–672. [[CrossRef](#)] [[PubMed](#)]
99. Vondrak, T.; Ávila Robledillo, L.; Novák, P.; Koblížková, A.; Neumann, P.; Macas, J. Characterization of Repeat Arrays in Ultra-Long Nanopore Reads Reveals Frequent Origin of Satellite DNA from Retrotransposon-Derived Tandem Repeats. *Plant J.* **2020**, *101*, 484–500. [[CrossRef](#)] [[PubMed](#)]
100. Han, M.-J.; Shen, Y.-H.; Gao, Y.-H.; Chen, L.-Y.; Xiang, Z.-H.; Zhang, Z. Burst Expansion, Distribution and Diversification of MITEs in the Silkworm Genome. *BMC Genom.* **2010**, *11*, 520. [[CrossRef](#)]
101. Deprá, M.; Ludwig, A.; Valente, V.L.; Loreto, E.L. Mar, a MITE Family of HAT Transposons in Drosophila. *Mob. DNA* **2012**, *3*, 13. [[CrossRef](#)]
102. Lu, C.; Chen, J.; Zhang, Y.; Hu, Q.; Su, W.; Kuang, H. Miniature Inverted-Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in Oryza Sativa. *Mol. Biol. Evol.* **2012**, *29*, 1005–1017. [[CrossRef](#)]
103. Kapitonov, V.V.; Jurka, J. Helitrons on a Roll: Eukaryotic Rolling-Circle Transposons. *Trends Genet.* **2007**, *23*, 521–529. [[CrossRef](#)]
104. Bennetzen, J.L.; Wang, H. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annu. Rev. Plant Biol.* **2014**, *65*, 505–530. [[CrossRef](#)] [[PubMed](#)]
105. Wawrzyniak, P.; Płucienniczak, G.; Bartosik, D. The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. *Front. Microbiol.* **2017**, *8*, 2353. [[CrossRef](#)] [[PubMed](#)]
106. Wong, L.H.; Choo, K.H.A. Evolutionary Dynamics of Transposable Elements at the Centromere. *Trends Genet.* **2004**, *20*, 611–616. [[CrossRef](#)]
107. Garrido-Ramos, M. Satellite DNA: An Evolving Topic. *Genes* **2017**, *8*, 230. [[CrossRef](#)]
108. Thakur, J.; Packiaraj, J.; Henikoff, S. Sequence, Chromatin and Evolution of Satellite DNA. *Int. J. Mol. Sci.* **2021**, *22*, 4309. [[CrossRef](#)] [[PubMed](#)]
109. Perea-Resa, C.; Blower, M.D. Centromere Biology: Transcription Goes on Stage. *Mol. Cell Biol.* **2018**, *38*, e00263-18. [[CrossRef](#)] [[PubMed](#)]
110. Nambiar, M.; Chuang, Y.-C.; Smith, G.R. Distributing Meiotic Crossovers for Optimal Fertility and Evolution. *DNA Repair* **2019**, *81*, 102648. [[CrossRef](#)] [[PubMed](#)]
111. Talbert, P.B.; Henikoff, S. What Makes a Centromere? *Exp. Cell Res.* **2020**, *389*, 111895. [[CrossRef](#)]
112. Gržan, T.; Despot-Slade, E.; Meštović, N.; Plohl, M.; Mravinac, B. CenH3 Distribution Reveals Extended Centromeres in the Model Beetle Tribolium Castaneum. *PLoS Genet.* **2020**, *16*, e1009115. [[CrossRef](#)]
113. Amorim, I.C.; Sotero-Caio, C.G.; Costa, R.G.C.; Xavier, C.; de Moura, R.d.C. Comprehensive Mapping of Transposable Elements Reveals Distinct Patterns of Element Accumulation on Chromosomes of Wild Beetles. *Chromosome Res.* **2021**, *29*, 203–218. [[CrossRef](#)]
114. Jiang, J.; Birchler, J.A.; Parrott, W.A.; Kelly Dawe, R. A Molecular View of Plant Centromeres. *Trends Plant Sci.* **2003**, *8*, 570–575. [[CrossRef](#)] [[PubMed](#)]
115. Mroczek, R.J.; Dawe, R.K. Distribution of Retroelements in Centromeres and Neocentromeres of Maize. *Genetics* **2003**, *165*, 809–819. [[CrossRef](#)] [[PubMed](#)]
116. Guo, X.; Su, H.; Shi, Q.; Fu, S.; Wang, J.; Zhang, X.; Hu, Z.; Han, F. De Novo Centromere Formation and Centromeric Sequence Expansion in Wheat and Its Wide Hybrids. *PLoS Genet.* **2016**, *12*, e1005997. [[CrossRef](#)] [[PubMed](#)]
117. Presting, G.G. Centromeric Retrotransposons and Centromere Function. *Curr. Opin. Genet. Dev.* **2018**, *49*, 79–84. [[CrossRef](#)] [[PubMed](#)]
118. Leo, L.; Marchetti, M.; Giunta, S.; Fanti, L. Epigenetics as an Evolutionary Tool for Centromere Flexibility. *Genes* **2020**, *11*, 809. [[CrossRef](#)] [[PubMed](#)]

119. Shang, W.-H.; Hori, T.; Martins, N.M.C.; Toyoda, A.; Misu, S.; Monma, N.; Hiratani, I.; Maeshima, K.; Ikeo, K.; Fujiyama, A.; et al. Chromosome Engineering Allows the Efficient Isolation of Vertebrate Neocentromeres. *Dev. Cell* **2013**, *24*, 635–648. [[CrossRef](#)] [[PubMed](#)]
120. Gao, D.; Jiang, N.; Wing, R.A.; Jiang, J.; Jackson, S.A. Transposons Play an Important Role in the Evolution and Diversification of Centromeres among Closely Related Species. *Front. Plant Sci.* **2015**, *6*, 216. [[CrossRef](#)]
121. Chang, C.-H.; Chavan, A.; Palladino, J.; Wei, X.; Martins, N.M.C.; Santinello, B.; Chen, C.-C.; Erceg, J.; Beliveau, B.J.; Wu, C.-T.; et al. Islands of Retroelements Are Major Components of Drosophila Centromeres. *PLoS Biol.* **2019**, *17*, e3000241. [[CrossRef](#)]
122. Wolfgruber, T.K.; Sharma, A.; Schneider, K.L.; Albert, P.S.; Koo, D.-H.; Shi, J.; Gao, Z.; Han, F.; Lee, H.; Xu, R.; et al. Maize Centromere Structure and Evolution: Sequence Analysis of Centromeres 2 and 5 Reveals Dynamic Loci Shaped Primarily by Retrotransposons. *PLoS Genet.* **2009**, *5*, e1000743. [[CrossRef](#)]
123. Zhang, H.-H.; Shen, Y.-H.; Xu, H.-E.; Liang, H.-Y.; Han, M.-J.; Zhang, Z. A Novel HAT Element in Bombyx Mori and Rhodnius Prolixus: Its Relationship with Miniature Inverted Repeat Transposable Elements (MITEs) and Horizontal Transfer. *Insect Mol. Biol.* **2013**, *22*, 584–596. [[CrossRef](#)]
124. Sharma, A.; Presting, G.G. Evolution of Centromeric Retrotransposons in Grasses. *Genome Biol. Evol.* **2014**, *6*, 1335–1352. [[CrossRef](#)] [[PubMed](#)]
125. Maheshwari, S.; Ishii, T.; Brown, C.T.; Houben, A.; Comai, L. Centromere Location in Arabidopsis Is Unaltered by Extreme Divergence in CENH3 Protein Sequence. *Genome Res.* **2017**, *27*, 471–478. [[CrossRef](#)] [[PubMed](#)]
126. Satović, E.; Vojvoda Zeljko, T.; Luchetti, A.; Mantovani, B.; Plohl, M. Adjacent Sequences Disclose Potential for Intra-Genomic Dispersal of Satellite DNA Repeats and Suggest a Complex Network with Transposable Elements. *BMC Genom.* **2016**, *17*, 997. [[CrossRef](#)] [[PubMed](#)]
127. Shang, W.-H.; Hori, T.; Toyoda, A.; Kato, J.; Popendorf, K.; Sakakibara, Y.; Fujiyama, A.; Fukagawa, T. Chickens Possess Centromeres with Both Extended Tandem Repeats and Short Non-Tandem-Repetitive Sequences. *Genome Res.* **2010**, *20*, 1219–1228. [[CrossRef](#)]
128. Wang, L.; Wang, Y.; Zhang, J.; Feng, Y.; Chen, Q.; Liu, Z.-S.; Liu, C.-L.; He, W.; Wang, H.; Yang, S.-F.; et al. Comparative Analysis of Transposable Elements and the Identification of Candidate Centromeric Elements in the Prunus Subgenus Cerasus and Its Relatives. *Genes* **2022**, *13*, 641. [[CrossRef](#)]
129. Saint-Leandre, B.; Nguyen, S.C.; Levine, M.T. Diversification and Collapse of a Telomere Elongation Mechanism. *Genome Res.* **2019**, *29*, 920–931. [[CrossRef](#)]
130. George, J.A.; DeBaryshe, P.G.; Traverse, K.L.; Celniker, S.E.; Pardue, M.-L. Genomic Organization of the Drosophila Telomere Retrotransposable Elements. *Genome Res.* **2006**, *16*, 1231–1240. [[CrossRef](#)]
131. Dover, G. Molecular Drive: A Cohesive Mode of Species Evolution. *Nature* **1982**, *299*, 111–117. [[CrossRef](#)]
132. Elliott, B.; Richardson, C.; Jasin, M. Chromosomal Translocation Mechanisms at Intronic Alu Elements in Mammalian Cells. *Mol. Cell* **2005**, *17*, 885–894. [[CrossRef](#)]
133. Hedges, D.J.; Deininger, P.L. Inviting Instability: Transposable Elements, Double-Strand Breaks, and the Maintenance of Genome Integrity. *Mutat. Res.* **2007**, *616*, 46–59. [[CrossRef](#)]
134. Chen, J.; Huang, Q.; Gao, D.; Wang, J.; Lang, Y.; Liu, T.; Li, B.; Bai, Z.; Luis Goicoechea, J.; Liang, C.; et al. Whole-Genome Sequencing of Oryza Brachyantha Reveals Mechanisms Underlying Oryza Genome Evolution. *Nat. Commun.* **2013**, *4*, 1595. [[CrossRef](#)] [[PubMed](#)]
135. Gilbert, D.M.; Bridges, M.C.; Strother, A.E.; Burckhalter, C.E.; Burnette, J.M.; Hancock, C.N. Precise Repair of MPing Excision Sites Is Facilitated by Target Site Duplication Derived Microhomology. *Mob. DNA* **2015**, *6*, 15. [[CrossRef](#)] [[PubMed](#)]
136. Rodgers, K.; McVey, M. Error-Prone Repair of DNA Double-Strand Breaks. *J. Cell Physiol.* **2016**, *231*, 15–24. [[CrossRef](#)]
137. Chang, H.H.Y.; Pannunzio, N.R.; Adachi, N.; Lieber, M.R. Non-Homologous DNA End Joining and Alternative Pathways to Double-Strand Break Repair. *Nat. Rev. Mol. Cell. Biol.* **2017**, *18*, 495–506. [[CrossRef](#)]
138. Liu, P.; Lacaria, M.; Zhang, F.; Withers, M.; Hastings, P.J.; Lupski, J.R. Frequency of Nonallelic Homologous Recombination Is Correlated with Length of Homology: Evidence That Ectopic Synapsis Precedes Ectopic Crossing-Over. *Am. J. Hum. Genet.* **2011**, *89*, 580–588. [[CrossRef](#)] [[PubMed](#)]
139. Parks, M.M.; Lawrence, C.E.; Raphael, B.J. Detecting Non-Allelic Homologous Recombination from High-Throughput Sequencing Data. *Genome Biol.* **2015**, *16*, 72. [[CrossRef](#)] [[PubMed](#)]
140. Al-Zain, A.M.; Symington, L.S. The Dark Side of Homology-Directed Repair. *DNA Repair* **2021**, *106*, 103181. [[CrossRef](#)] [[PubMed](#)]
141. Morris, R.T.; Drouin, G. Ectopic Gene Conversions in the Genome of Ten *Hemiascomycete* Yeast Species. *Int. J. Evol. Biol.* **2010**, *2011*, 970768. [[CrossRef](#)]
142. Mazón, G.; Lam, A.F.; Ho, C.K.; Kupiec, M.; Symington, L.S. The Rad1–Rad10 Nuclease Promotes Chromosome Translocations between Dispersed Repeats. *Nat. Struct. Mol. Biol.* **2012**, *19*, 964–971. [[CrossRef](#)]
143. Marsano, R.M.; Milano, R.; Minervini, C.; Moschetti, R.; Caggese, C.; Barsanti, P.; Caizzi, R. Organization and Possible Origin of the Bari-1 Cluster in the Heterochromatic H39 Region of Drosophila Melanogaster. *Genetica* **2003**, *117*, 281–289. [[CrossRef](#)]

### **3.2. CAPÍTULO 2:**

**ELEMENTOS DE TRANSPOSIÇÃO COMO RESERVATÓRIOS DE  
UNIDADES REPETITIVAS DURANTE O SURGIMENTO DE UM NOVO DNA  
SATÉLITE: O CASO DOS ELEMENTOS hAT NO GENOMA DE *P. carvalhoi* E  
DE *X. tropicalis***



**ELEMENTOS DE TRANSPOSIÇÃO COMO RESERVATÓRIOS DE  
UNIDADES REPETITIVAS DURANTE O SURGIMENTO DE UM NOVO DNA  
SATÉLITE: O CASO DOS ELEMENTOS hAT NO GENOMA DE *P. carvalhoi* E  
DE *X. tropicalis***

### 3.2.1 CONTEXTUALIZAÇÃO

A descrição dos TEs no genoma de anfíbios ainda está restrita ao estudo do conteúdo de TEs (=composição e fração dos elementos em relação ao genoma completo) inferidos por meio da anotação dos genomas recém montados durante seus depósito em bancos de dados públicos, com poucos trabalhos aprofundando-se na caracterização molecular (Sinzelle et al., 2005; Roger et al., 2014; Gazolla et al., 2022), ou na sua evolução de um determinado grupo (Gama et al., 2022, Zuo et al., 2023) e em seu mapeamento nos cariótipos (Nascimento et al., 2015). Baseados no fluxograma metodológico aplicados tradicionalmente em outros grupos taxonômicos de vertebrados e invertebrados (Casavant et al., 2000; Valente et al., 2011; Nascimento et al., 2015; Noronha et al., 2016), o isolamento de TEs nos genomas tem se baseado no uso de primers heterólogos como forma inicial de se obter as sequências de TEs. Essa premissa tem sentido devido à grande conservação encontrada nos motivos genéticos das enzimas envolvidas na mobilização desses elementos (por exemplo, gene da transposase, transcriptase reversa) mesmo em linhagens com alto grau de divergência evolutiva.

O relato do estudo de caso descrito a seguir teve início em achados encontrados inesperadamente utilizando esse mesmo fluxograma metodológico no genoma de *P. carvalhoi*. Assim, neste capítulo exploramos esses achados em *P. carvalhoi* e expandimos para *Xenopus tropicalis*, uma espécie filogeneticamente relacionada e modelo para estudos de genômica comparativa.

#### 3.2.1.1 - Caso: Cópias derivadas do elemento hAT no genoma de *P. carvalhoi*

Estudando o conteúdo de DNA repetitivo no genoma de *Pipa carvalhoi* por meio de isolamento de sequências repetitivas via reação da polimerase em cadeia (PCR, em inglês) e mapeamento cromossômico por meio de experimentos de FISH de sequências repetitivas (gene da Histona H3, por exemplo), Zattera (2019) realizou uma série de tentativas de amplificação de sequências de TEs utilizando *primers* heterólogos.

A sequência que exploramos neste capítulo foi isolada utilizando *primers* heterólogos originalmente desenvolvidos para amplificar o retrotransposon Rex3 do peixe *Xiphophorus maculatus* (Volf et al., 2000; Volf et al. 1999; Volf et al., 2001). Este conjunto de primers já teve seu uso de forma heteróloga para uma gama de espécies de peixes (Ozouf-Costaz et al., 2004; Teixeira et al., 2009; Silva et al., 2013; Costa et al., 2016; Prizon et al., 2018; Traldi et al., 2019).

Interessantemente, o trabalho de Nascimento et al. (2015) obteve sucesso no isolamento de outro retroelemento, o Rex 1, nos genomas de anfíbios de espécies do gênero *Physalaemus* utilizando o conjunto de primers também descrito por Volf et al. (2001). Os achados destes autores nos instigaram a verificar o potencial de amplificação deste e de outros retroelementos previamente isolados de peixes, como Rex 1, Rex 3, Rex 5 e Rex 6. Apesar de todas estas subfamílias de retrotransposons receberem denominações semelhantes [Rex1 (Retroelement of *Xiphophorus* 1), Rex3 (Retroelement of *Xiphophorus* 3), Rex5 (Retroelement of *Xiphophorus* 5) e Rex6 (Retroelement of *Xiphophorus* 6) (Volf et al., 2001, Volf et al., 2000, Volf et al., 1999)], eles não se tratam de um grupo parálogo (Carducci et al., 2018; Gama et al., 2022), mas sim, clados distintos dentre a árvore evolutiva dos retroelementos da ordem LINE.

Obtivemos sucesso na amplificação de sequências a partir do uso do conjunto de primers para amplificação do retroelemento da superfamília Rex 3. A seguir descrevemos as metodologias aplicadas e os desdobramentos destes achados.

### 3.2.1.2 Metodologias iniciais de isolamento das sequências

#### *Amostras biológicas e obtenção de suspensões celulares*

Os espécimes de *Pipa carvalhoi* utilizados neste estudo foram coletados no município de Buíque, estado de Pernambuco, Brasil, sob autorização do SISBIO/Instituto Chico Mendes de Conservação da Biodiversidade (protocolo 55481-1). Após análise, os espécimes foram depositados no Museu de História Natural da Universidade Federal de Alagoas (MHN-UFAL).

Para a obtenção de metáfases mitóticas, suspensões celulares do intestino foram tratadas com colchicina a 2% (na proporção de 0,02mL/g do animal), preparadas seguindo o protocolo de King e Rofe (1976) e Schmid (1978).

#### *Isolamento e clonagem*

Amostras de DNA genômico foram isoladas dos tecidos hepático e muscular de *P. carvalhoi* utilizando o método de extração de DNA TNES conforme descrito por Bruschi et al (2012). A amplificação por PCR foi realizada utilizando os primers REX3-F (5'-CGGTGAYAAAGGGCAGCCCTG-3') e REX3-R (5'-TGGCAGACNNGGGTGGTGGTG-3'), seguindo Volff et al (1999).

Um gradiente de temperatura variando de 54°C a 60°C foi utilizado para determinar a temperatura ótima de anelamento. O programa de amplificação consistiu em uma etapa inicial de desnaturação a 95°C por 5 minutos, seguida de 35 ciclos a 94°C por 1 minuto, anelamento a 58°C - 60°C por 1 minuto, extensão a 72°C por 1 minuto e 30 segundos e etapa final de extensão a 72°C por 5 minutos.

O produto amplificado da PCR foi submetido à eletroforese em gel de agarose (1,0%), e a banda obtida foi extraída do gel utilizando o EasyPure Quick Gel Extraction Kit (PROMEGA) seguindo o protocolo do fabricante. O fragmento de DNA purificado foi ligado ao vetor pJET 1.2/blunt e transformado na bactéria *Escherichia coli* da cepa TOP10 usando o CloneJET PCR Cloning Kit (Thermo Scientific) de acordo com as

instruções do fabricante. O DNA plasmidial dos clones selecionados foi extraído pelo método de mini-prep de Sambrook et al (1989), e os clones foram submetidos ao sequenciamento de Sanger.

#### *Hibridação in situ fluorescente (FISH)*

O clone Pcarvalhoi\_clone17 foi randomicamente utilizado para gerar a sonda, marcada por PCR com 11-digoxigenina-dUTP. Para reduzir a hibridização inespecífica, a sonda foi precipitada juntamente com 1ng/mL de DNA de esperma de salmão como DNA competidor, e o tempo de hibridização das lâminas foi de 72 horas. O protocolo de FISH seguiu o método de Traut et al (2001), e as preparações cromossômicas foram examinadas utilizando-se um microscópio confocal Nikon, com posterior montagem das metáfases com o uso do Adobe Photoshop.

#### *Análise de sequência*

Para o sequenciamento bidirecional dos clones foi aplicado o método de Sanger com o kit Big Dye Terminator, e as sequências resultantes foram editadas manualmente através do programa BioEdit para posterior análise. As *reads* de *P. carvalhoi* foram utilizadas como *query* para realizar o BLAST local (*tBLASTn*), contra o genoma de *Pipa carvalhoi*, a fim de recuperar os *contigs* com a sequência do elemento isolado. Estes contigs foram recuperados do genoma completo usando códigos Python, escritos para filtrar os resultados e acessar as sequências de interesse conforme descritos em Gama et al (2022).

As sequências editadas foram comparadas quanto à identidade e similaridade com sequências e genes de outros organismos usando o software *Basic Local Alignment Search Tool* (BLAST), disponível no site do National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/blast/>).

Os contigs isolados do genoma de *P. carvalhoi* foram analisados utilizando o banco de dados REPBASE (disponível em: <https://www.girinst.org/repbase/>). As buscas

por domínios conservados de proteínas foram realizadas por meio da ferramenta de busca online de CD-search (disponível em: <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). A confirmação da presença de TIRs foi realizada através do BLAST online, onde as cópias foram analisadas individualmente usando Blastn, com a opção "align two or more sequences" e "optimize for somewhat similar sequences (Blastn)".

### **3.2.2. RESULTADOS INICIAIS E ESTRATÉGIAS PARA SUA COMPREENSÃO**

Os fragmentos obtidos das amostras de *P. carvalhoi* apresentaram duas bandas distintas, uma com 250 pb e outra com 300 pb, na temperatura de anelamento de 58°C. Aproximadamente 50 clones foram obtidos da banda mais leve (~250pb), dos quais 15 foram enviados para sequenciamento. Paralelamente, realizamos o mapeamento cromossômico de um desses clones (Pcarvalhoi\_clone 17) por ensaios de FISH.

Os experimentos de FISH usando a sonda derivada do fragmento Pcarvalhoi\_clone17 revelaram sinais de hibridização clusterizados nos dois homólogos dos cromossomos dos pares 1,2, 4, 5, 6, 7, 8 e 9 do cariótipo de *P. carvalhoi*. A maioria dos sinais foi observado nas regiões pericentroméricas dos pares homólogo 2 e 4, exibindo marcações em seus braços curtos, juntamente com marcações leves, porém específicas, nos pares 5, 6 e 7. Além disso, os sinais de hibridização foram identificados em 1q, 8q e como um bloco subterminal correspondente ao local NOR no homólogo do par 9q (Figura 1).

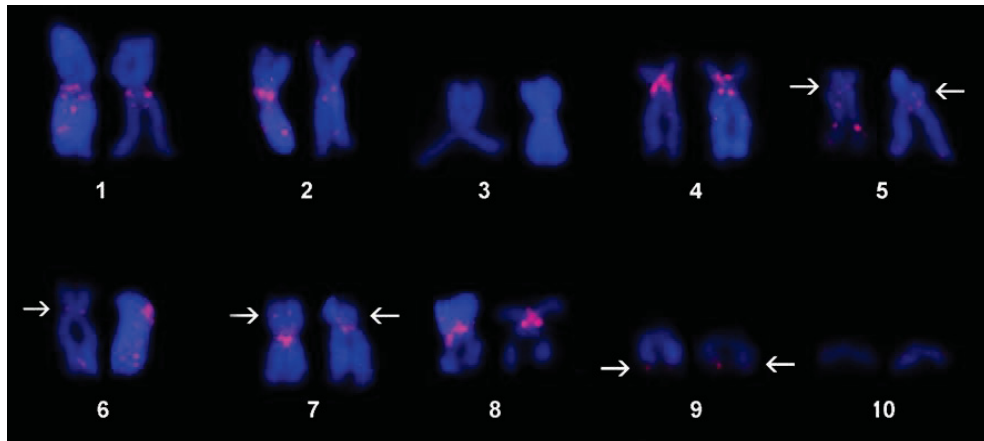


Figura 1: Cromossomos metafásicos de *P. carvalhoi* submetidos à FISH com sonda isolada a partir dos primers de Rex 3. Fonte: Autor (2023).

Utilizando a sequência *Pcarvalhoi\_clone17* como query em buscas no genoma sequenciado de *P. carvalhoi*, realizamos buscas intragenômicas a fim de verificar a densidade de cópias com alta similaridade com a sequência de nosso clone, a fim de ter maior representatividade da sequência isolada, já que a grande maioria do sequenciamento dos clones não obteve resultados satisfatórios. Os contigs foram selecionados com base nos parâmetros de melhor confiança (E-value) e bit-score, e a análise considerou a porcentagem de similaridade das cópias com o clone.

Vários contigs com altos escores foram encontrados, mostrando entre 91-94% de identidade com a sequência isolada do clone. Com a seleção dos melhores contigs recuperados, conduzimos uma análise de Máxima Verossimilianção (ML= Maximum Likelihood) para estimar uma árvore de sequências para esta matriz. O perfil recuperado para nossa árvore, não enraizada, demonstra o alto grau de conservação entre as sequências recuperadas no genoma, conforme observado na Figura 2.

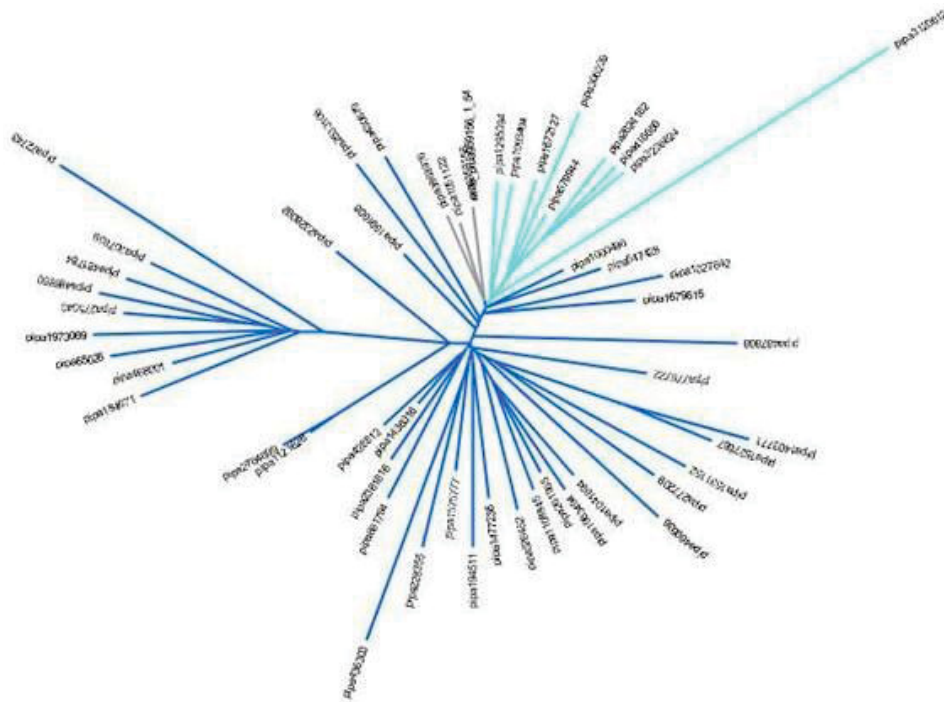


Figura 2: Dendrograma das sequências de hAT inferida pelo critério de ML (Maximum Likelihood) a partir das cópias com maior nível de similaridade com a sequência Pcarvalhoi\_clone17 recuperadas do genoma de *P. carvalhoi*.

Nossas análises iniciais de bioinformática nos surpreenderam ao encontrar similaridade da sequência Pcarvalhoi\_clone17 com cópias do transposon da superfamília hAT, um elemento de classe II (transposon de DNA) presente no genoma de *Xenopus tropicalis*. O elemento em questão, se encontra depositado no REPBASE sob denominação hAT-XT, com uma série de depósitos de famílias numeradas sequencialmente, como elementos das famílias hAT-10\_XT, hAT-9\_XT e hAT-9A\_XT. Nossos dados recuperaram similaridade entre as cópias isoladas do genoma de *P. carvalhoi* em relação aos elementos hAT-10\_XT, hAT-9\_XT e hAT-9A\_XT variando de 65,19% a 72,73% (Tabela 1). A identidade entre os contigs e as sequências isoladas dos elementos de *X. tropicalis*, segundo o BLAST, variou entre 68,25% e 94,44% (Tabela 1).

Tabela 1: Análise comparativa da identidade do clone isolado do genoma de *P. carvalhoi* (Pcarvalhoi\_clone17) em relação aos contigs recuperados do genoma da referida espécie (colunas 1 e 2) e sua comparação quanto à similaridade (%) do melhor hit recuperado no REPBASE (colunas 4 e 5). Valores de identidade (%) foram recuperados de comparações realizadas no BLAST. A análise de presença e tamanho (pb) de repetições invertidas terminais (TIRs) também é fornecida. XT: *Xenopus tropicalis*; (-): ausência de TIRs.

<b>Contigs</b>	<b>Identidade (%)</b>	<b>Mismatches</b>	<b>Rebase</b>	<b>Similaridade (%)</b>	<b>Identidade (%)</b>	<b>Potenciais TIRs (?)</b>
<i>pipa1000496</i>	94.37	8	hAT_10 XT	65,26	81,82	19 pb
<i>pipa959156</i>	93.16	11	hAT_9 XT	71,49	71,05	19 pb
<i>pipa579944</i>	93.16	11	hAT_9 A_XT	66,81	68,14	19-17 pb
<i>pipa3863976</i>	93.04	11	hAT_9 A_XT	0,6685	86,96	19 pb
<i>pipa2073077</i>	92.64	8	hAT_9 XT	0,6519	68,25	-
<i>pipa2034876</i>	92.31	13	hAT_10 XT	0,7273	73,33	-
<i>pipa1672127</i>	92.31	13	hAT_9 A_XT	0,6831	75,61	-
<i>pipa1370773</i>	92.37	11	hAT_9 XT	0,7048	94,44	-
<i>pipa2100846</i>	91.88	14	hAT_10 XT	0,6883	81,58	-
<i>pipa2091853</i>	91.88	14	hAT_10 XT	0,6757	81,58	-
<i>pipa1651950</i>	91.88	14	hAT_9 A_XT	0,6864	75,51	-
<i>pipa1438316</i>	91.88	14	hAT_9 A_XT	0,6681	81,4	-

Analisando a estrutura molecular das cópias recuperadas nos melhores contigs isolados e da sequência do clone, não foram encontrados domínios conservados em nenhum dos contigs analisados, mesmo que pequenas ORFs altamente degeneradas tenham sido identificadas. No entanto, repetições terminais invertidas (TIRs) e um sítio alvo de duplicação (TSD) foram identificados em quatro contigs (*pipa100049*, *pipa959156*, *pipa579944* e *pipa3863976*) e na sequência clonada (Figura 3). Tanto no clone quanto nestes contigs, observou-se uma sequência de 19 pb (5' TCCTGTCCAAGTTGCTGGT 3') compatível com TIR, conforme tamanho e características já relatados para esses elementos estruturais na superfamília hAT (Wicker 2007).

Não foi possível identificar os TIRs nos demais contigs. Duas hipóteses foram levantadas: (i) que possivelmente devido a estas cópias estarem localizados no final dos contigs (*pipa2073077*, *pipa1672127*, *pipa2091853*, *pipa1651950*); ou (ii) por serem cópias muito pequenas do elemento (*pipa2034876*, *pipa1370773*, *pipa2100846*), e acabam mascarando a presença destes elementos estruturais de hAT. Dois contigs (*pipa100049*, *pipa959156*) e a sequência clonada apresentaram TSDs, correspondendo a 8 pb na região a montante (5'GCACCGTA 3') da TIR na porção 5', e 8 pb na porção 3'



(5' GCACGGTC 3') da TIR. Ambas as cópias apresentaram dois polimorfismos de nucleotídeo único (SNPs) nas TSDs da porção 3' do fragmento [(CG); (AC)], apesar de serem cópias distintas, como evidenciado pela pequena variação nucleotídica entre elas (Figura 3).

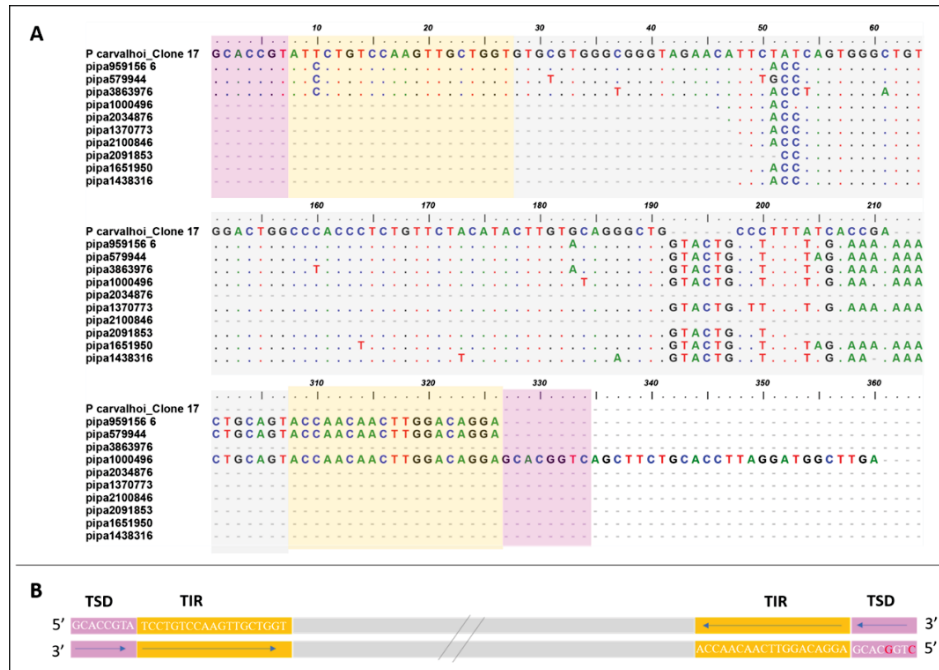


Figura 3: Estrutura molecular predita para as cópias onde foram recuperadas potenciais TIRs

Assim, ausência de ORFs conservadas e presença de potenciais TIRs nos levantaram uma hipótese de que a sequência do clone isolado poderia se tratar de um MITE (do inglês, *miniature inverted-repeat transposable elements*). Essas sequências repetitivas que ocorrem em uma ampla gama de organismos, incluindo plantas, animais e fungos (Ye et al., 2016), são versões curtas (variando entre 100 a 800 pb) truncadas e não-autônomas de elementos transponíveis autônomos pertencentes ao tipo classe II de TEs (Venkatesh e Nandini, 2020). Por serem sequências que não apresentam o domínio da transposase, estas dependem de outras proteínas codificadas por elementos transponíveis autônomos para sua mobilização para novas posições no genoma (Wicker, 2007). Ainda, é interessante notar que MITEs estão frequentemente localizados em regiões eucromáticas e que são, portanto, ricas em genes (Han et al., 2013; Chen et al., 2013).

As sequências MITE apresentam outra característica interessante, conhecida como os eventos de *burst*. Devido ao seu pequeno tamanho, essas repetições conseguem escapar dos mecanismos de silenciamento do genoma (Chen et al., 2019), permitindo sua amplificação acúmulo em regiões cromossômicas específica (*burst events*), potencialmente facilitando o surgimento e o estabelecimento de um novo DNA satélite (Han et al., 2010; Deprá et al., 2012; Scalvenzi e Pollet, 2014). Isso poderia explicar a organização clusterizada nas regiões pericentroméricas dos cromossomos de *P. carvalhoi*.

Em anfíbios, dois casos são interessantes de serem observados dentro deste contexto e nos levaram a buscar avaliar melhor esta hipótese, já que ambos foram reportados para espécies de *Xenopus*, grupo-irmão de *P. carvalhoi*. Em *Xenopus laevis*, a sequência repetitiva em tandem Xstir originou-se por meio da duplicação da repetição interna do elemento transponível Xmix MITE (Hikosaka, 2004), ao passo que em *Xenopus tropicalis* o satDNA miDNA4 foi gerado a partir da amplificação de repetições tandem do MITE TC1-mariner, com rápida homogeneização das matrizes repetitivas e, ao longo do tempo, a geração do satDNA (Scalvenzi & Pollet, 2014).

Desta forma, construímos estratégias de análises bioinformáticas que pudessem nos dar evidências para a avaliação destas hipóteses. A seguir, exploramos melhor essas questões:

*1. Se a sequência Pcarvalhoi\_clone17 se trata de um MITE de um elemento hAT, ainda é possível encontrar cópias íntegras do elemento original no genoma de P. carvalhoi?*

Para avaliar esta questão, conduzimos uma busca para estabelecer qual das famílias de hAT-XT estariam presentes no genoma de *P. carvalhoi* e qual a cópia com melhor identidade e maior integridade. Desta forma, foi conduzida uma análise de tblastn das ORFs associadas aos elementos hAT\_9 e hAT\_10 de *Xenopus tropicalis* em relação ao genoma de *P. carvalhoi*.

Para a realização deste teste, foram reconstruídas cópias destes elementos a partir de fragmentos específicos das ORFs de *Xenopus tropicalis*, além de regiões flanqueadoras de 10 kb para cada lado. As sequências obtidas mostraram-se menores que o ideal, apresentando cerca de 13-14kb dentro da espécie *P. carvalhoi* e é interessante notar que a maioria dos matches foram localizados em contigs menores de 5 kb, como indicado na Figura 4. Considerando que a parte codificadora de proteínas (ORF) do hAT\_10 possui 3.4 kb, os dados demonstram que a sequência que encontramos é significativamente menor do que a ORF esperada para esse elemento.

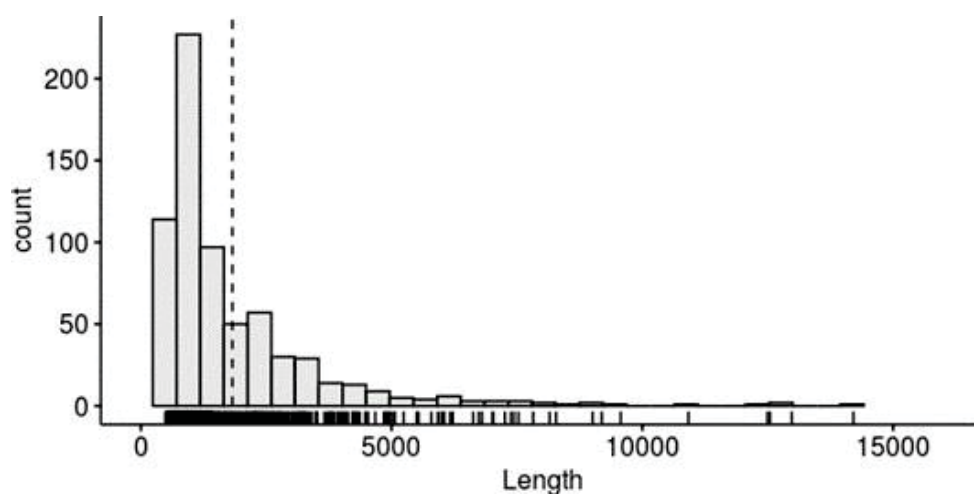


Figura 4. Histograma com os resultados do blast da open read frame (ORF) de hAT\_10 no genoma de *Pipa carvalhoi*. É possível observar que a maior densidade de matches estão em contigs menores de 5kb, marcados como blocos em preto em nosso esquema.

A mesma metodologia de extração e ampliação das sequências foi utilizada para comparação através do blast das ORFs de hAT\_9 e hAT\_10 depositados no REPBASE com as cópias que podem ser encontradas na versão completa do genoma *Xenopus tropicalis*. As cópias obtidas foram consideravelmente maiores do que as obtidas em *P. carvalhoi*, apresentando cerca de 20 kb, evidenciando cópias menos fragmentadas e sofrendo menores efeitos da montagem do genoma. Porém, a alta fragmentação do elemento fica evidente com a predição das ORFs, uma vez que a maior densidade delas apresenta menos de 2 kb (Figura 5).

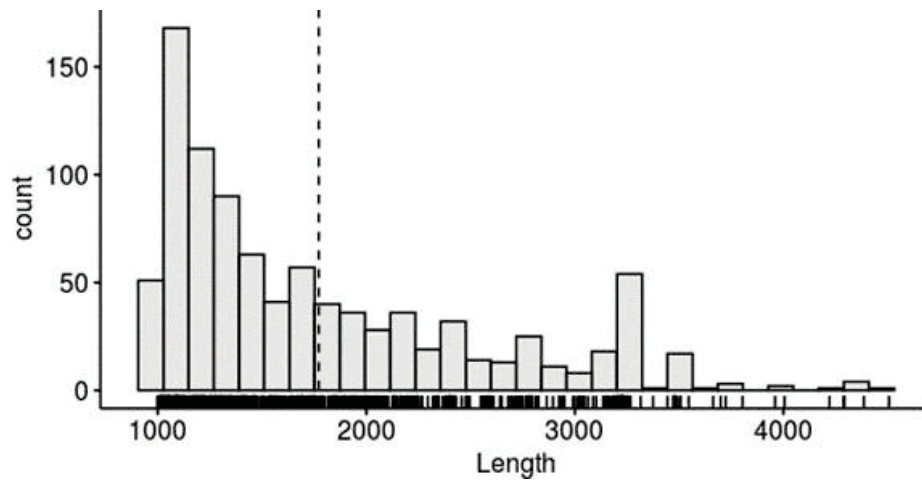


Figura 5. Histograma com os resultados do blast do open read frame (ORF) de hAT\_10 no genoma de *Xenopus tropicalis*. É possível observar que a maior densidade dos matches estão em contigs com menos de 2kb.

Os resultados dessas análises indicam que a sequência Pcarvalhoi\_clone17 apresenta uma maior similaridade com o elemento hAT\_10, com uma quantidade significativamente maior de correspondências quando comparado ao hAT\_9. Apesar destes esforços das buscas, ainda restava a dúvida se realmente existiam sequências íntegras deste elemento no genoma de *P. carvalhoi*, uma vez que a versão de montagem do genoma de *P. carvalhoi* utilizada nestas análises ainda está bastante fragmentada. Por isso, tentamos a estratégia de uma montagem “de novo” dos TEs com base nas reads brutas do genoma.

Para esta estratégia, utilizamos a ferramenta DNAPipeTE, na qual tem incorporada em seu pipeline a ferramenta Trinity, ideal para este tipo de abordagem. Assim, a caracterização do conteúdo completo de TEs que compõem o genoma de *P. carvalhoi* foi obtida a partir dos dados de Illumina raw short-reads utilizando do DNAPipeTE v.1.4c (Goubert et al., 2015). Essa pipeline foi desenvolvida para encontrar, classificar e quantificar TEs e outros elementos repetitivos em dados de NGS de baixa cobertura (<1X) sem a necessidade de um genoma montado. O DNAPipeTE foi utilizado nos dados brutos do genoma de *P. carvalhoi*, sequenciado na plataforma Illumina HiSeq-400, 46x de cobertura. Os dados brutos das

reads foram trimados utilizando Trimmomatic (Bolger et al., 2014) e estão disponíveis em [http://200.236.3.93/xeno/velvet/contigs\\_k105.fa](http://200.236.3.93/xeno/velvet/contigs_k105.fa).

Os resultados indicaram que 10% do genoma era formado por transposons de DNA (Figura 6). Em termos de proporção de contigs montados, os matches com o hAT\_10 de *Xenopus tropicalis* representavam apenas 370 dos 95 mil contigs de TEs montados (Figura 6). Contigs similares ao hAT\_9 ocorreram com uma frequência significativamente menor (apenas 34 contigs), indicando que o hAT\_9 deve ser escasso no genoma de *Pipa carvalhoi*. Essa abordagem confirmou nossas análises por blast, porém, também não conseguimos recuperar cópias íntegras de hAT-10 no genoma de *P. carvalhoi*.

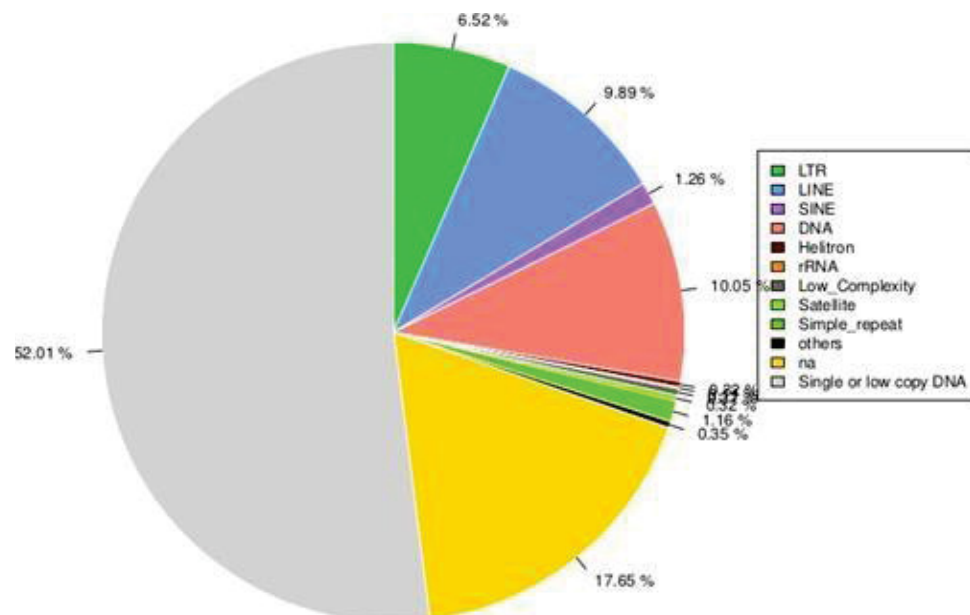


Figura 6. Gráfico obtido pela ferramenta DNAPipeTE com o conteúdo de elementos transponíveis no genoma de *Pipa carvalhoi*. Os resultados demonstram que cerca de 10% do genoma é composto por transposons de DNA, classe que inclui o elemento hAT.

De fato, nossas buscas não encontraram cópias íntegras deste elemento no genoma de *P. carvalhoi*, porém, vale destacar que, a família hAT-10XT depositada no REPBASE se trata de uma única sequência consenso, e assim, restava a dúvida se este elemento se tratava de uma cópia íntegra do genoma de *X. tropicalis*.

1. Como seriam as cópias íntegras de *hAT\_10* no genoma de *P. carvalhoi* e *X. tropicalis*? Estratégias para obtenção de uma cópia consenso e compreensão da estrutura molecular do elemento *hAT*.

Quando isolamos a cópia completa do elemento *hAT\_10* de *X. tropicalis* e o maior contig do genoma de *P. carvalhoi* (contig *pipa481784*), algo nos surpreendeu na estrutura molecular atípica destas cópias: em ambos os elementos foi possível observar a presença de um tipo de repetições internas repetidas “in tandem” dentro da estrutura destes elementos. Curiosamente, as unidades monoméricas têm tamanhos e sequências distintas entre os dois elementos, mas se repetem em sua forma de organização (Figura 7).

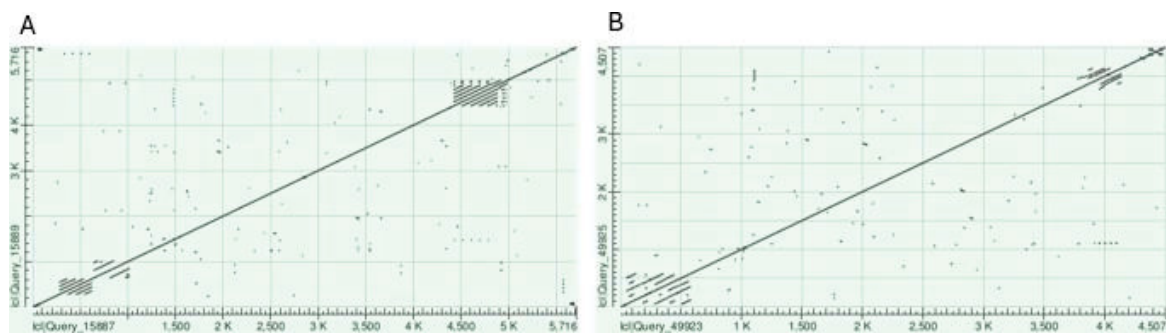


Figura 7: Análise da distribuição dos fragmentos encontrados através do blast do elemento *hAT\_10* no genoma de (A) *Xenopus tropicalis* (cópia consenso REPBASE) e (B) *Pipa carvalhoi* (contig *pipa481784*). A área pontilhada (vermelha) representa unidades repetitivas observadas dentro do elemento. Retângulo representa posição que a sequência *Pcarvalhoi\_clone17* alinha na respectiva cópia.

Em *hAT-10* de *X. tropicalis*, a primeira repetição representa uma unidade monomérica constituída por 90pb, variando no número de repetições (de 6 a 8 repetições in tandem) dependendo da cópia analisada e a segunda repetição, de 99 pb, que se repete de 12 a 20 vezes. Já na cópia de *Pipa carvalhoi* (contig *pipa481784*), trata-se uma unidade monomérica de 46 pb, estando presente em 18 repetições sequenciais no contig analisado. A segunda unidade monomérica não foi possível de ser estabelecida em virtude do alto grau de fragmentação para esta região.

O clone isolado está em uma região mais central do elemento, e apesar de não ter recuperado ORFs conservadas, foi possível observar seu alinhamento na região mais interna do elemento, na região das ORFs dessas cópias.

Não existem até o momento relatos para organizações moleculares semelhantes a estas em elementos da superfamília hAT. Na verdade, a presença de unidades monoméricas internas na estrutura do TE praticamente não foram relatadas na literatura. Para nosso melhor conhecimento, recentemente Zeljko et al. (2020) reportaram a presença de unidades monoméricas semelhantes à DNA satélite em elementos Helentron de *Crassostrea gigas*. Neste caso, ao analisar as sequências flanqueadoras de 13 clusters de satDNA, foi observado que as matrizes mais abundantes estavam integradas como repetições centrais de elementos não-autônomos Cg\_HINE (Helentons). Essas repetições em tandem localizadas centralmente foram, também, encontradas na forma de satDNAs clássicos, dando força à hipótese destes TEs no surgimento e expansão destes satélites.

Desta forma, compreender se esta organização de repetições internas do elemento hAT de *Pipa carvalhoi* estaria sempre associada ao elemento completo ou com suas ORFs das cópias incompletas passou a ser determinante para compreendermos se a unidade monomérica é parte do elemento e estaria gerando um novo DNA satélite ou se este elemento estaria apenas espalhando cópias de um DNA satélite já presente neste genoma. Ademais, precisávamos entender se as repetições internas presentes na cópia consenso de hAT\_10XT em *X. tropicalis* faria realmente a parte do elemento ou seria apenas um artefato de montagem. Assim, para resolver essas questões, foi necessário o estabelecimento de diferentes estratégias de testes que pudessem fornecer evidências para a compreensão desta estrutura. Todos os testes são descritos a seguir e foram conduzidos para os genomas de *P. carvalhoi* e *X. tropicalis*.

**Teste 1: Comparação da distribuição de cópias dos hATs nos genomas de *Pipa carvalhoi* e *Xenopus tropicalis***

Ao realizar a análise de cópias e cobertura do consenso do hAt\_10 contra o genoma de *Xenopus tropicalis*, foi possível verificar a presença de cerca de 11 mil cópias fragmentadas (Figura 8).

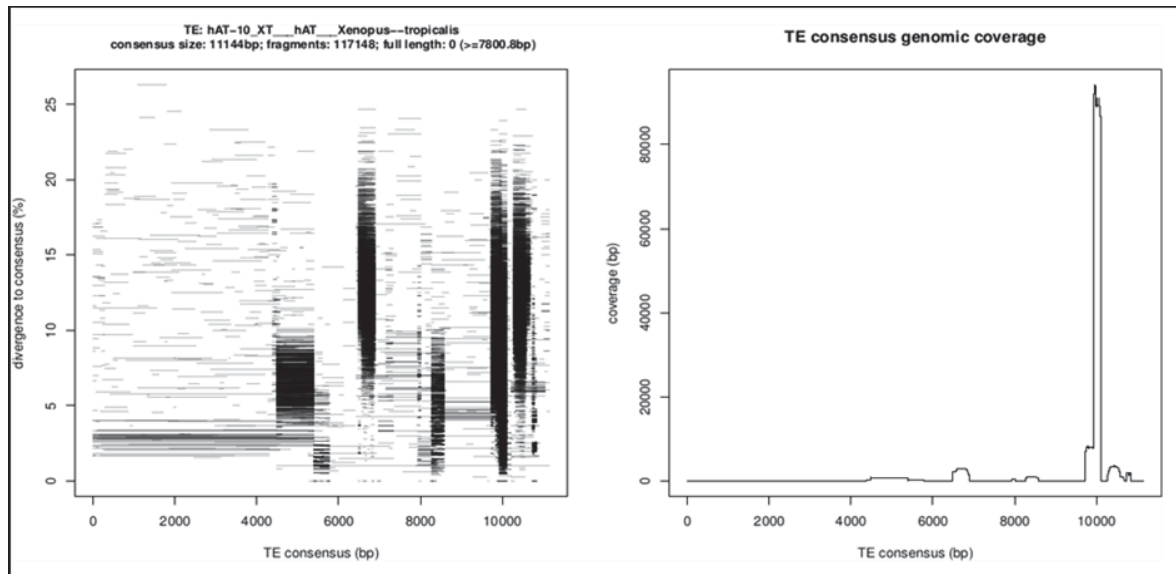


Figura 8. Análise de cópias e cobertura do consenso do hAt\_10 contra o genoma de *Xenopus tropicalis*, evidenciando a fragmentação e divergência das cópias em relação ao consenso do elemento transponível.

Porém, também foi possível observar que grande parte desses fragmentos, representa a região de repetições diretas presentes no consenso e a distribuição dos fragmentos também se apresentou de maneira desigual havendo maior número de cópias dos satélites presentes na região de 10kb, parte compõe a região de repetições diretas do hAT\_10 (Figura 9).



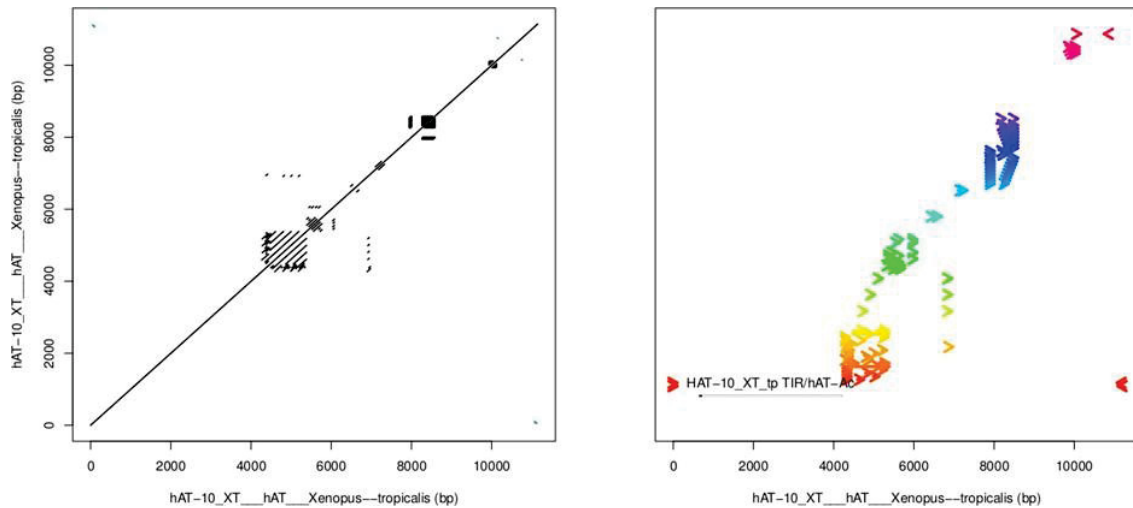


Figura 9. Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Xenopus tropicalis*, onde é possível perceber um grande número de cópias repetitivas na região entre 4000-6000 pares de base, bem como na porção de 10.000 pares de base.

Quando consideramos a mesma análise do contig pipa481784 contra o genoma de *P. carvalhoi*, foi possível perceber que grande parte das cópias deste elemento estão fragmentadas, com exceção de uma, que abrange toda a região de ORF de hAT. Há, ainda, a ausência de cobertura de cópias em uma região próxima de 4.1kb, podendo indicar o fim do elemento (Figura 10) e, justamente nesta região se encontra o intervalo entre o grupo de satélites e o conjunto de repetições diretas e invertidas presentes no final do contig. Além disso, o número de fragmentos dos satélites da região 4kb é muito similar a cobertura da região do domínio proteico do hAT (Figura 11). Essa região genômica é bastante desafiadora, uma vez que, se as repetições monoméricas internas estão organizadas em tandem, as falhas de montagem destes contigs tem um efeito ainda maior sob essas regiões.

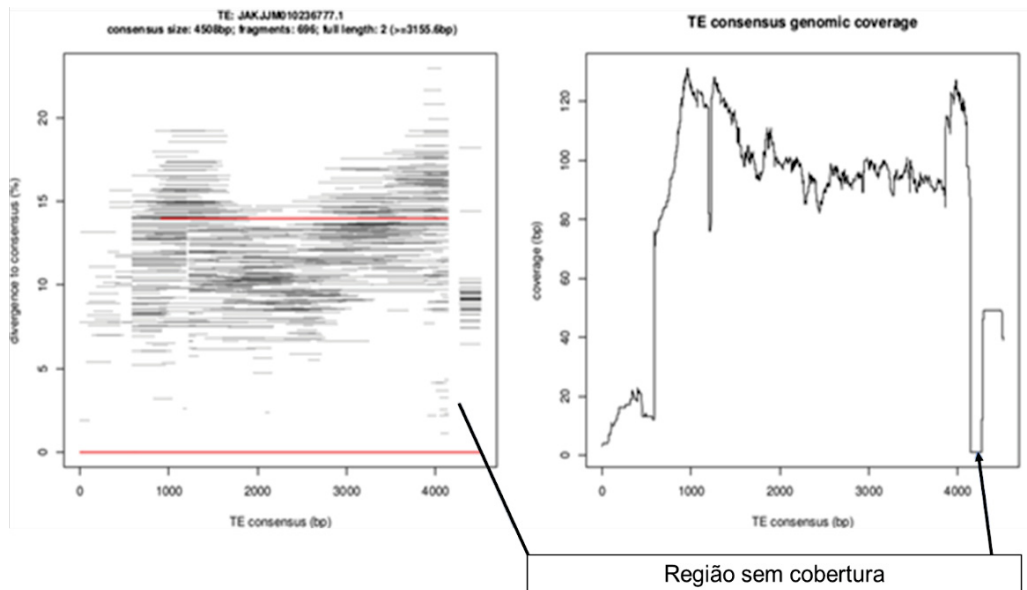


Figura 10. Análise de cópias e cobertura do consenso do hAT\_10 contra o genoma de *Pipa carvalhoi*, onde é possível observar que a maior parte das cópias deste elemento estão fragmentadas, mas uma delas, evidenciada pela linha em vermelho, abrange toda a região da ORF do hAT. A ausência de cobertura na região por volta de 4,1 kb parece indicar o fim da cópia do elemento de transposição no contig pipa481784.

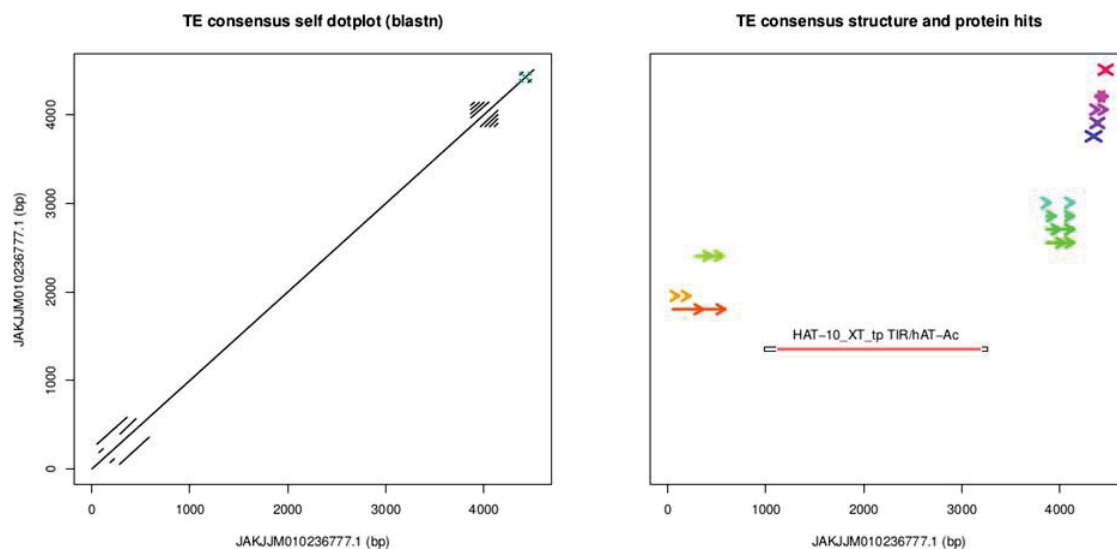


Figura 11. Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Pipa carvalhoi*. É possível observar que a cobertura das repetições do início do contig é bem menor quando comparada a região de 4kb do elemento.

Diante da fragmentação das sequências obtidas e considerando que esta análise foi realizada com base em uma única cópia do clone, a abordagem foi direcionada para a construção de uma sequência consenso do elemento hAT\_10. Essa construção teve como base a estrutura do contig pipa481784.

## Teste 2: Construção de uma sequência consenso com base em contig pipa481784

Devido à alta fragmentação do genoma de *P. carvalhoi*, foram utilizadas ferramentas como RepeatMasker, RepeatModeler e scripts auxiliares, para a construção deste consenso. É importante destacar que para estas análises foi pressuposto que o elemento real de hAT tenha a mesma estrutura do contig pipa481784. Essa premissa foi assumida arbitrariamente para que pudéssemos dar sequências em nossas reconstruções.

O consenso obtido proporcionou uma ideia detalhada da estrutura do elemento hAT\_10, aumentando a resolução na região de repetições no início da sequência. Além disso, revelou a presença de mais repetições na região de 4 kb em comparação à análise realizada apenas com a cópia de pipa481784. Entretanto, a região após 4,2 kb passou a apresentar um reduzido número de correspondências, indicando potencialmente que ela não faz parte do elemento transponível ou temos um problema de representatividade de sequências nesta região do alinhamento que deve ser explorado melhor nos testes seguintes (Figura 12).

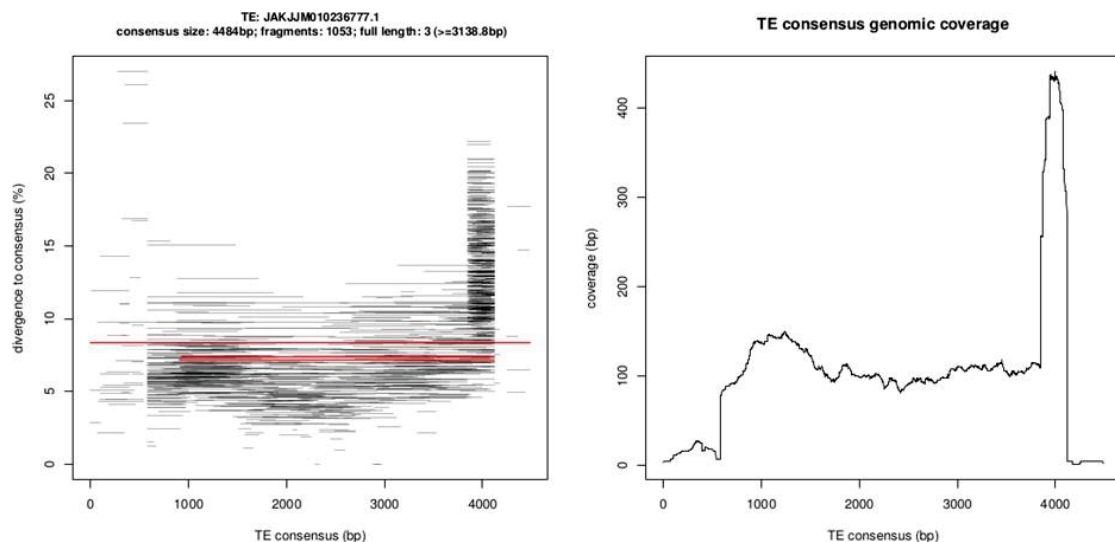


Figura 12. Análise de cópias e cobertura do consenso do hAT\_10 contra o genoma de *Pipa carvalhoi*. No consenso fica evidenciada a resolução da região de repetições do início da sequência, além da presença de mais repetições na região de 4kb.

O consenso revelou a presença de dois potenciais "MITES" dentro da sequência, identificados por sequências de repetição terminal inversa (TIR) flanqueando uma região não codificante (Figura 13). Uma das sequências apresenta 316 pares de bases e está localizado nas regiões 1633-1651 a 1948-1930, enquanto o segundo apresenta um tamanho de 1302 pares de base, presente nas regiões 1541-1556 a 2842-2827.

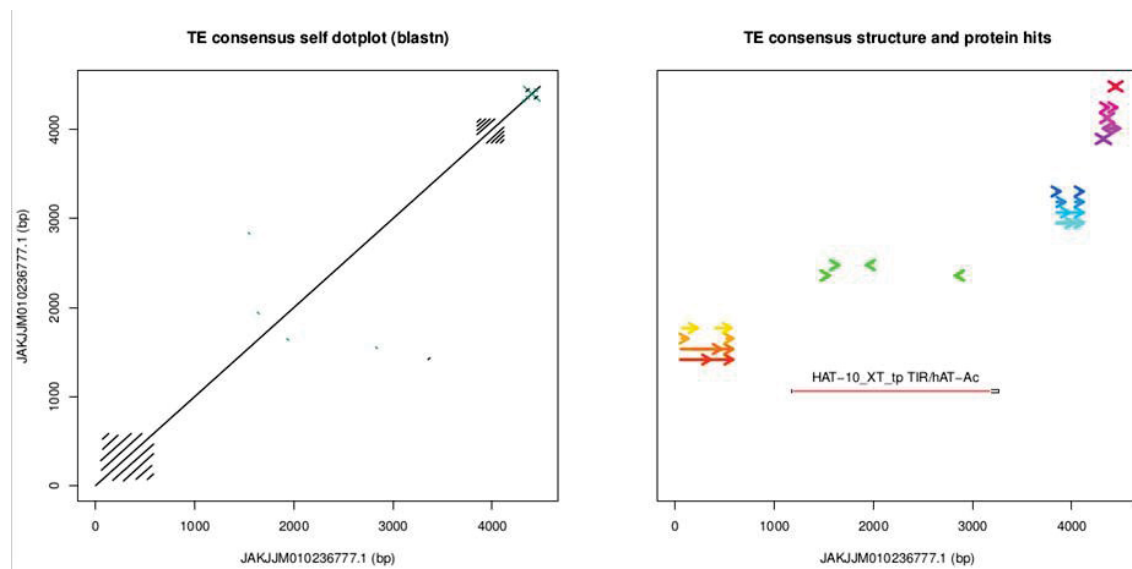


Figura 13. Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Pipa carvalhoi*. A presença de potenciais MITES está evidenciada pelas setas verdes, presentes na região da ORF do elemento de hAT\_10.

A análise da ORF do consenso confirmou que se tratava do elemento hAT\_10, embora este consenso ainda não tenha apresentado uma ORF íntegra, uma vez que o domínio Dimer não está presente (Figura 14), o que nos levou a uma tentativa de reconstrução do elemento através dos domínios conservados deste transposon.

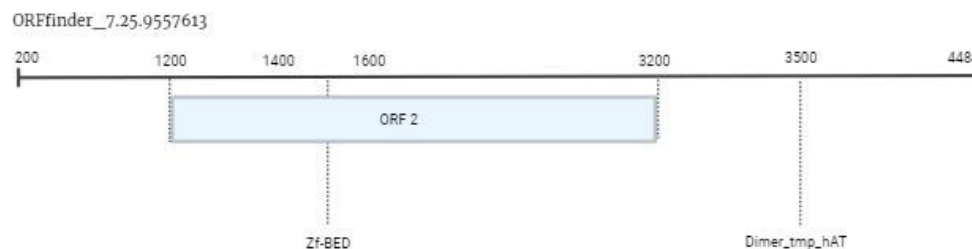


Figura 14. Desenho esquemático da análise da ORF do elemento de hAT através da ferramenta ORFfinder, destacando que a sequência se trata de um hAT\_10 incompleto, uma vez que domínio Dimer ainda não está na ORF, ainda que o domínio de dedo de zinco (Zf-BED) esteja dentro da ORF.

**Teste 3: Reconstrução do consenso a partir do domínio conservado de hAT presente no contig pipa481784.**

Ao contrário do teste 2, onde se utilizou toda a sequência pipa481784, aqui o foco foi na região codificante do hAT e parte do domínio Dimer, que se apresentou fora da ORF na análise anterior. Utilizou-se a região codificante do hAT como input para um blastn, onde foram selecionados HSPs (*high scoring pairs*) com até 25% de divergência, e cobertura em relação à query de pelo menos 30%, resultando em 97 cópias. Estas cópias, sem região flanqueadora, foram alinhadas com o Refiner para obter uma ORF íntegra que contivesse os dois domínios do hAT (Figura 15).

O consenso resultante dessa abordagem foi comparado com a cópia pipa481784, mostrando uma identidade de 90%. Porém, para que a obtenção da ORF estivesse completa, se fez necessário a extração da região flanqueadora, exigindo uma etapa de extensão do tamanho do consenso obtido e estudo das porções flanqueadoras.

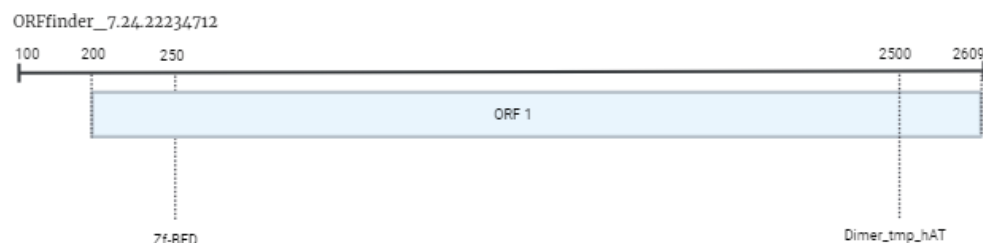


Figura 15. Desenho esquemático da análise da ORF do elemento de hAT através da ferramenta ORFfinder, destacando que a sequência se trata de um hAT<sub>10</sub>, destacando a obtenção da ORF completa.

**Teste 4: Extensão do consenso obtido para definição das regiões flanqueadoras do TE**

Este teste buscou estender o consenso obtido no teste 3 para definir as bordas do elemento hAT<sub>10</sub>. Se para o teste 3 foi pressuposto que a estrutura do TE se parecia com a estrutura do contig de *P. carvalhoi*, o presente teste utilizou de outra

abordagem, uma vez que pelo menos uma das bordas do contig de *P. carvalhoi* pareciam não fazer parte do consenso. Para isso, foram realizados alinhamentos transitivos par-a-par, manualmente inspecionados, das 97 cópias obtidas no teste 3 para reconstrução do alinhamento múltiplo. A matriz de Kimura 2-Parâmetros (K2P) foi ajustada para este alinhamento, buscando encontrar a ORF mais intacta possível. O consenso obtido com este teste foi consideravelmente maior (3294 pares de base) quando comparada à do teste 3 (2616 pares de base), estando muito mais próxima do valor da ORF de hAt\_10 de *X. tropicalis* (Figura 16).

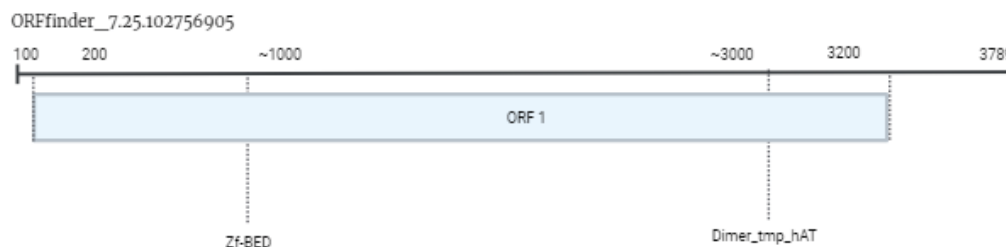


Figura 16. Desenho esquemático da análise da ORF consenso do elemento de hAT através da ferramenta ORFfinder. Com esta montagem, a ORF consenso apresentou um tamanho aproximado de 3,2 kb, sendo próxima ao tamanho da ORF original de *Xenopus tropicalis*.

**A região repetitiva que estava presente na região 4 kb de pipa481784 realmente faz parte do elemento**, porém, a região inicial e final no contig pipa481784 não gerou um bom suporte no alinhamento, apresentando a ocorrência de muitas bases discordantes nessas regiões impossibilitando a extensão do consenso. Dentro das análises ainda foram identificadas as regiões que se assemelham a MITEs dentro da ORF (Figura 17), caracterizadas por sequências de repetição terminal inversa (TIR) flanqueando uma região não codificante.

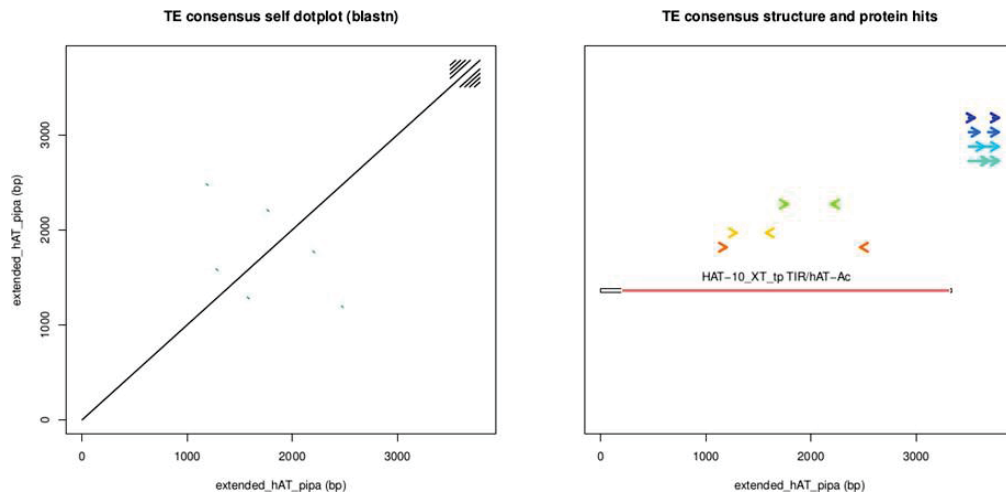


Figura 17. Análise da distribuição dos fragmentos encontrados através do blast do elemento de hAT\_10 no genoma de *Pipa carvalhoi*. É possível observar que, com a sequência consenso de hAT, as repetições encontradas na região da ORF do elemento, bem como na região de 4kb, permanecem sendo mapeadas no genoma da espécie.

### ***Teste 5: Uma investigação sobre os supostos MITEs encontrado***

Visando ampliar a sequência consenso obtida anteriormente para verificar melhor os possíveis MITEs, a sequência consenso foi estendida a partir de cópias mais fragmentadas e divergentes que se alinhavam nas bordas do consenso prévio, com posterior refinamento para que o melhor modelo possível fosse gerado. A região central apresenta uma reconstrução mais segura quando comparada aos primeiros e últimos 500 nucleotídeos, devido à alta fragmentação presente nesta região. Essa fragmentação pode ser tanto degeneração do elemento transponível como fragmentação na montagem genômica.

Focando nas análises dos potenciais MITEs, primeiro foi analisado o MITE 1, com tamanho de aproximadamente 1302 nt, localizado nas regiões 1898-1913 a 3199. Foi possível perceber que o número de cópias que se alinham exclusivamente com estes potenciais elementos foi limitado, com os hits se estendendo para as regiões flangeadoras em 5'. Embora seja interessante destacar que o que pode indicar uma associação frequente dessas regiões com o restante do elemento de hAT (Figura 18).

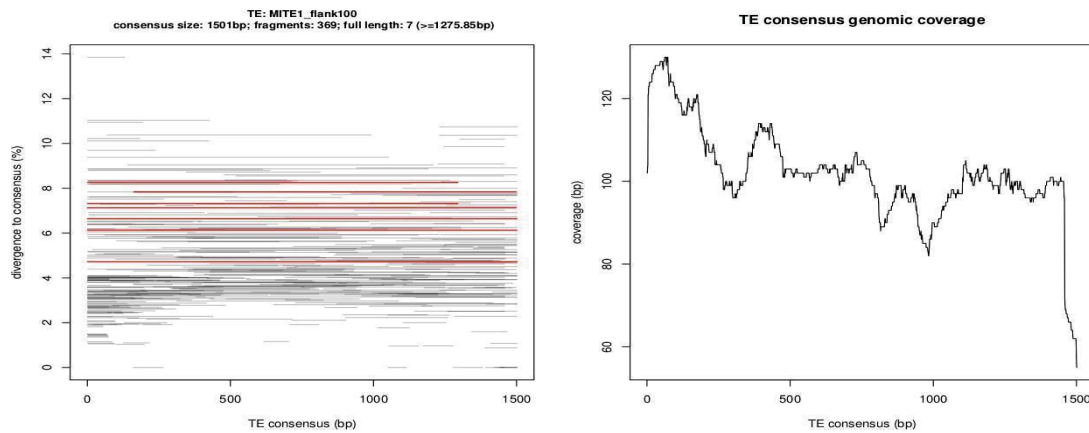


Figura 18. Análise de cópias e cobertura do possível MITE 1. As primeiras e as últimas 100 bases são da região flanqueadora do potencial MITE. As linhas em vermelho representam os fragmentos que alinharam com esta região.

O segundo MITE apresentava tamanho de 316 nucleotídeos e se localizava na região 1990-2008 a 2305-2287. Para esta repetição observa-se que, embora existam fragmentos consideráveis alinhados com a região flanqueadora, a maioria das cópias contínua para as regiões flanqueadoras 3' e 5', sugerindo uma forte ligação contínua com o restante do hAT, em vez de uma unidade de replicação independente (Figura 19).

**Estas análises indicam que, por não se tratar de unidades de replicação independentes, as sequências observadas dentro do elemento de hAT não se tratam de MITEs, como se acreditava anteriormente.**

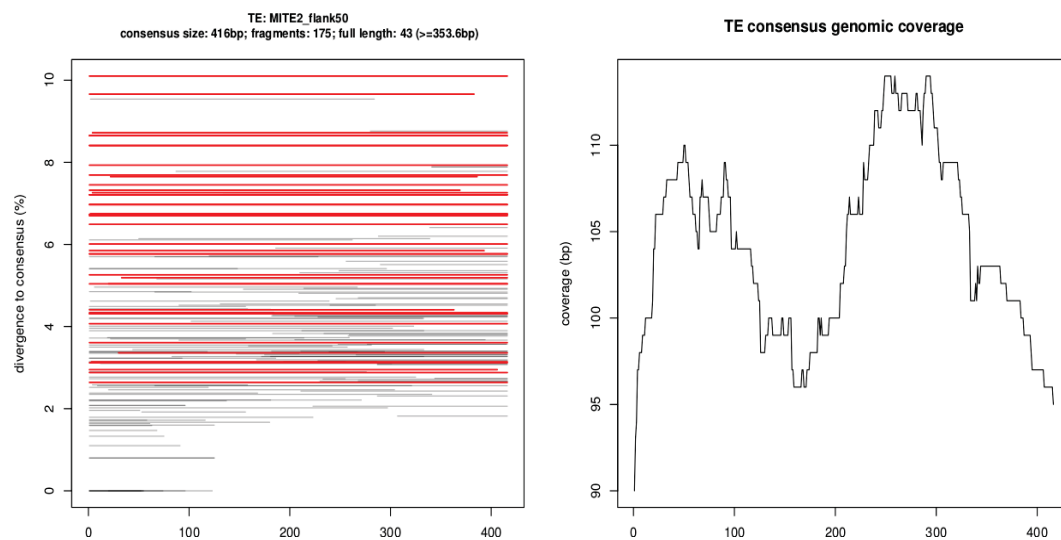


Figura 19: Análise de cópias e cobertura do possível MITE 2. As primeiras e as últimas 50 bases são da região flanqueadora do potencial MITE. As linhas em vermelho representam os fragmentos que alinharam com esta região.



### **Impressões e hipóteses levantadas a partir destas reconstruções:**

(i) Os elementos hAT-10 não se encontram íntegros no genoma de *P. carvalhoi* e nem em *X. tropicalis*.

O fato de não encontrarmos essas cópias íntegras pode nos sugerir que esses elementos podem estar se degenerando e como resultado desta degeneração, deixando as cópias dos elementos repetitivos internos nos genomas. Essas sequências então, no ciclo de vida deste transposon, poderiam estar atuando como substrato para emergência de um novo DNA satélite (Zattera and Bruschi, 2022). O alto grau de conservação entre as cópias destas unidades monoméricas poderia indicar que possa estar ocorrendo mecanismos de homogeneização intragenômica, compatíveis com o modelo de evolução em concerto (Concerted evolution process, para revisão ver Elder e Turner, 1995).

(ii) A sequência Pcarvalhoi\_clone17 não se trata de MITE no genoma de *P. carvalhoi*, mas está organizada de maneira repetitiva?

Não encontramos evidências de que a sequência Pcarvalhoi\_clone17 se trate de um MITE no genoma de *P. carvalhoi*, porém, observamos uma organização repetitiva no genoma por meio de experimentos de FISH. Nossa hipótese é a de que, uma vez que a sequência Pcarvalhoi\_clone17 compõe parte interna das ORFs de hAT-10, o acúmulo de transposons degenerados em uma dada região cromossômica poderia favorecer a detecção desta sequência nos experimentos de FISH. De fato, quando inspecionamos mais contigs do genoma de *P. carvalhoi*, com menor E-value, conseguimos recuperar contigs com mais de um alinhamento com a região do clone, porém bastante degenerado. A inspeção destes demais contigs podem demonstrar essa organização agrupada das cópias degeneradas de hAT-10 em sítios cromossômicos próximos.

(iii) As unidades repetitivas fazem parte do elemento hAT-10 no genoma de *P. carvalhoi* e em *X. tropicalis*;

Sempre que encontramos as cópias das unidades repetitivas monoméricas, elas estavam associadas às ORFs do elemento hAT-10, sem observar cópias genômicas isoladas em outras porções genômicas. Isso nos leva a crer que a degeneração deste elemento transposon possa estar realmente gerando um novo DNA satélite. Neste nosso modelo, a homogeneização das cópias de unidades monoméricas internas por meio de crossing-over desigual ou conversão gênica poderiam estar favorecendo a aproximação das cópias desta unidade repetitiva em uma organização agrupada. Este modelo precisará de testes específicos para sua futura validação, porém, este caso parece se tratar de algo distinto do que se verificou por Zeljko et al. (2020). Esses autores reportaram a presença de unidades monoméricas semelhantes à DNA satélite em elementos Helentron de *Crassostrea gigas*. No entanto, neste caso, os autores também encontraram essas sequências satDNAs em outras regiões genômicas, sem associação com o TE, dando força à hipótese de que, neste caso, os TEs estão envolvidos na expansão destes satélites.

(iv) A estrutura molecular composta por unidades repetitivas internas trata-se de uma característica estrutural dos elementos da superfamília hAT no genoma de Anuros? Seria o elemento hAT\_10 de anfíbios um reservatório para novos DNAs satélites?

Com esses resultados aparentemente inéditos, realizamos uma última análise destinada a compreender se esta estrutura molecular composta por unidades monoméricas internas nesses TEs (DNA satélite-like) se trata de uma característica desta família no genoma de anfíbios. Para isso, utilizamos os consensos gerados para *X. tropicalis* e *P. carvalhoi* como query em buscas por homologia (blastn) no genoma de espécies de anuros que apresentam seu genoma sequenciado, montado e disponível para acesso.

Cópias homólogas ao transposon hAT\_10 foram identificadas em mais doze espécies de anuros (Tabela 2). Um dos resultados que chamam a atenção se refere à

espécie de *Gastrophryne carolinensis*, onde mais de 40 mil cópias deste consenso foram identificadas.

Tabela 2. Tabela com resultados das análises de blastn com base no consenso final de hAT\_10 gerado e os genomas de anuros já sequenciados, com foto no número de cópias desta repetição encontrados em cada genoma.

<b>Espécie</b>	<b>Genoma</b>	<b>Número de Cópias</b>
<i>Pyxicephalus adspersus</i>	GCA_04786255.1_Pads_1.0	1398
<i>Bufo gargarizans</i>	GCA_014858855.1_ASM1485885v1	28976
<i>Engystomops pustulosus</i>	GCA_019512145.1_UCB_Epus_1.0	6149
<i>Pipa parva</i>	GCA_019650415.1_UCB_Ppar_1.0	89
<i>Pipa carvalhoi</i>	GCA_021901965.1-W530	9
<i>Phyllomedusa bahiana</i>	GCA_025379985.1_CIBIO_Pbah_1.0	1335
<i>Gastrophryne carolinensis</i>	GCA_027917425.1_aGascar1.pri	43349
<i>Hyla sarda</i>	GCA_029499605.1_aHylSar1.hap1	21214
<i>Rhinella marina</i>	GCA_900303285.1-RM170330	754
<i>Bufo bufo</i>	GCA_905171765.1-aBufBuf1.1	32866
<i>Leptodactylus fallax</i>	GCA_947044405.1_aLepFal1.1	5960
<i>Staurois parvus</i>	GCA_95U30385.1_Staurois_parvus_ger	1388

Quanto às regiões de satélites localizadas após a ORF do elemento de hAT\_10, foi possível verificar a presença dessas repetições em diferentes espécies através de uso de blast por linha de comando. Alguns dados são interessantes de serem destacados como a espécie de *Pyxicephalus adspersus* (Tabela 3), onde a região de cobertura em relação ao consenso (91%) com 4448 pb, se encontra entre as posições 756 – 5018, sendo uma região semelhante à de satélite observada em *Pipa carvalhoi*. Ainda, foi possível observar que em *Bufo gargarizans* tal região acontece entre as posições 756 – 5048, e em *Phyllomedusa bahiana* entre 756 e 5415.

Esses resultados ressaltam que as repetições de satélites dentro deste elemento são comuns em diferentes espécies de anfíbios, incluindo aquelas filogeneticamente distantes. Esses dados são bastante promissores, visto o potencial desses elementos atuarem como reservatórios de monômeros para o surgimento de novas famílias de DNA satélite.

Tabela 3. Resultados obtidos pelo blastn comparando a sequência consenso obtida de hAT\_10 de *Pipa carvalhoi* com os genomas de anuros já sequenciados, buscando evidenciar as regiões de início e fim das repetições, destacando a cobertura em relação à query.

	query	alvo	identidade (%)	tamanho (pb)	inicio (nt)	final (nt)	cobertura da query (%)
<i>Pyxicephalus adspersus</i>	consenso_com_tirs_	CM016417.1	77.34	4448	756	5018	91
	consenso_com_tirs_	CM016417.1	76.30	4097	926	4855	91
	consenso_com_tirs_	CM016417.2	74.62	4424	823	5048	91
	consenso_com_tirs_	CM016417.3	77.69	3577	797	4312	91
	consenso_com_tirs_	CM016417.4	77.51	3673	756	4342	91
	consenso_com_tirs_	CM016417.5	76.55	3881	1321	5048	91
	consenso_com_tirs_	CM016417.6	77.40	3545	756	4228	91
	consenso_com_tirs_	CM016417.7	74.24	4368	830	5023	91
	consenso_com_tirs_	CM016417.8	78.69	3364	759	4057	91
	consenso_com_tirs_	CM016417.9	78.52	3273	1119	4309	91
<i>Bufo gargarizans</i>	consenso_com_tirs_	CM026473.1	77.53	4460	756	5048	91
	consenso_com_tirs_	CM026473.1	78.06	3970	756	4585	91
	consenso_com_tirs_	CM026473.1	76.54	4156	1001	4979	91
	consenso_com_tirs_	CM026473.1	77.08	3974	756	4585	91
	consenso_com_tirs_	CM026473.1	76.57	4161	991	4971	91
	consenso_com_tirs_	CM026473.1	77.14	3989	753	4647	91
	consenso_com_tirs_	CM026473.1	75.88	4187	1001	5016	91
	consenso_com_tirs_	CM026473.1	77.11	3227	756	4371	91
	consenso_com_tirs_	CM026473.1	75.48	4070	813	4748	91
	consenso_com_tirs_	CM026473.1	75.50	3934	756	4601	91
<i>Phyllomedusa bahiana</i>	consenso_com_tirs_	JAODAL010059185.1	82.40	4903	736	5415	86
	consenso_com_tirs_	JAODAL010059185.1	71.17	829	4216	4921	86
	consenso_com_tirs_	JAODAL010059185.1	70.55	859	4315	5048	86
	consenso_com_tirs_	JAODAL010059185.1	71.08	785	4218	4888	86
	consenso_com_tirs_	JAODAL010059185.1	71.84	696	4218	4818	86
	consenso_com_tirs_	JAODAL010059185.1	70.87	745	4407	5048	86
	consenso_com_tirs_	JAODAL010059185.1	71.23	716	4361	4972	86
	consenso_com_tirs_	JAODAL010059185.1	71.80	649	4216	4772	86
	consenso_com_tirs_	JAODAL010059185.1	71.84	593	4218	4726	86
	consenso_com_tirs_	JAODAL010059185.1	70.14	690	4453	5048	86

### 3.2.3. DISCUSSÃO

Os resultados obtidos em nossas tentativas de reconstrução destes elementos repetitivos apontam para um possível papel de reservatório de unidades monoméricas repetitivas dentro da estrutura molecular dos elementos hAT\_10 das espécies analisadas. Este achado inesperado e inédito aponta para o papel de TEs na remodelação dos genomas via interconversão de classes distintas de elementos repetitivos, servindo como um importante substrato molecular para a evolução genômica. A lista de produtos genômicos funcionais evolutivos emergentes a partir TEs é extensa e claramente subestimada (Trizzino et al., 2018). Além disso, novos elementos genômicos não funcionais também emergem dos TEs, como DNAs satélites e outros elementos repetitivos, que compõem a arquitetura cromossômica e participam do remodelamento

cariotípico durante a divergência evolutiva das linhagens (Paço et al., 2019; Zattera e Bruschi, 2022).

Por exemplo, hoje sabe-se que TEs são a fonte primária de inovações evolutivas em redes de regulação gênica (Lynch et al., 2011; Jacques et al., 2013; Sundaram et al., 2014; Quesneville, 2020), surgimento de novos genes (Long et al., 2003; Bennetzen, 2005; Kaessmann, 2010; Quesneville, 2020) ou em mecanismos complexos como a pluripotência celular (Macfarlan et al., 2012; Wang et al., 2020; Ma et al., 2022).

Os TEs exibem segmentos naturalmente repetitivos em sua estrutura, como Repetições Terminais Invertidas (TIRs), Repetições Terminais Longas (LTRs), e outras regiões não codificantes, que favorecem a ocorrência das interações inter e intracromossômicas. Eventos como (i) uma reação de transposição alternativa (Gray 2000), (ii) recombinação não-homóloga, (iii) *crossing-over* desigual ou (iv) reparo ectópico do DNA na interfase (Balachandran et al., 2022), podem gerar uma biblioteca de matrizes de repetição curtas derivadas de TEs que se transformam em novas unidades de monômeros que podem dar origem a novos DNAs repetitivos derivados de TEs (para revisão, ver Zattera e Bruschi, 2022).

Nomeada após a descrição de três famílias de transposons de DNA (*hobo*, *Activator* e *Tam3*), a superfamília hAT é uma das nove superfamílias da subclasse I de transposons de DNA (Handler & Gomez, 1997; Rubin et al, 2001; Wicker et al., 2007), e entre os organismos de anura com genoma sequenciado, elementos hAT já foram relatados em *Nanorana parkeri* (Sun et al., 2015), *Rhinella marina* (Edwards et al 2018), no qual o elemento *hAT-Ac* foi o mais abundante entre os transposons de DNA encontrados, e *Xenopus tropicalis* (Sun et al., 2015). Refinar as análises sobre a estrutura molecular desses elementos hAT são muito promissoras nos genomas de anfíbios, uma vez que, conforme nossos dados parecem indicar, esses elementos têm o potencial de causar interconversão de elementos repetitivos, contribuindo no surgimento de novos DNAs satélites nos genomas, por exemplo.

Considerando os resultados obtidos, essas repetições satélites estão presentes em espécies de anura mesmo filogeneticamente distantes, evidenciando a importância de estudos que complementem essa lacuna, assim como demais análises que possam comprovar e entender melhor a estrutura das repetições encontradas em *Pipa carvalhoi* e *Xenopus tropicalis*.

## **PERSPECTIVAS FUTURAS**

Visando ampliar os resultados obtidos e melhor compreender aqueles que já foram alcançados, será necessário compreender melhor o papel das sequências de hAT nos genomas de anuros. Para isso, considerando as espécies que apresentam seu genoma sequenciado, será analisada a qualidade destes genomas e a possibilidade da montagem de sequências consenso de hAT\_10 para cada destas espécies, através de curadoria manual. Através da montagem destes consensos, será possível identificar a presença ou ausência deste elemento, bem como se há a presença das repetições de satélite e as regiões cromossômicas onde elas ocorrem.

Ainda, o mapeamento cromossômico através de FISH destas sequências deverá ser realizado na espécie de *Xenopus tropicalis* para demonstrar as regiões cromossômicas onde o elemento de hAT\_10 é encontrado dentro da espécie.

Com estes resultados será possível compreender melhor a dinâmica deste elemento transponível em Anura, bem como a possível dispersão de sequências de satDNA que os elementos de hAT podem estar favorecendo dentro destes genomas.

## REFERÊNCIAS

- BALACHANDRAN, P. et al. 2022. Transposable element-mediated rearrangements are prevalent in human genomes. *Nature Communications*, v. 13, n. 1, p. 7115.
- BENNETZEN, J. L. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development*, v. 15, n. 6, p. 621–627.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120.
- BRUSCHI, D.P; BUSIN C.S; SIQUEIRA S; RECCO-PIMENTEL, S.M. 2012. Cytogenetic analysis of two species in the *Phyllomedusa hypochondrialis* group (Anura, Hylidae). *Hereditas*, v. 149, p. 34–40.
- CARDUCCI, F. et al. 2018. Rex retroelements and teleost genomes: An Overview. *International Journal of Molecular Sciences*, v. 19, p. 1–15.
- CASAVANT, N. C. et al. 2000. The end of the LINE?: lack of recent Ll activity in a group of south american rodents. *Genetics*, v. 154, n. 4, p. 1809–1817.
- CHEN, J. et al. 2019. Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice. *Nature Communications*, v. 10, n. 1.
- CHEN, J. et al. 2013. P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Research*, v. 42, n. D1, p. D1176–D1181.
- COATES, B. S. et al. 2010. A Helitron-Like transposon superfamily from lepidoptera disrupts (GAAA)<sub>n</sub> microsatellites and is responsible for flanking sequence similarity within a microsatellite family. *Journal of Molecular Evolution*, v. 70, n. 3, p. 275–288.
- COSTA, G. W. W. F. et al. 2016. The evolutionary dynamics of ribosomal genes, histone H3, and transposable Rex elements in the genome of atlantic snappers. *The Journal of Heredity*, v. 107, n. 2, p. 173–180.
- DEPRÁ, M. et al. 2012. Mar, a MITE family of hAT transposons in *Drosophila*. *Mobile DNA*, v. 3, n. 1.
- EDWARDS, R.J et al.2018. Draft genome assembly of the invasive cane toad, *Rhinella marina*. *Giga Science*.

- GAMA, J. M; et al. 2022. A genomic survey of LINE elements in Pipidae aquatic frogs shed light on Rex-elements evolution in these genomes. *Molecular Phylogenetics and Evolution*. Volume 168.
- GAZOLLA, C. B; et al. 2022. Evolutionary dynamics of DIRS-like and Ngaro-like retrotransposons in *Xenopus laevis* and *Xenopus tropicalis* genomes. *G3 Genes|Genomes|Genetics*, Volume 12, Issue 2.
- GENT, J. I.; DAWE, R. K. 2012. RNA as a structural and regulatory component of the centromere. *Annual Review of Genetics*, v. 46, n. 1, p. 443–453.
- GRAY, Y. H. M. 2000. It takes two transposons to tango. *Trends in Genetics*, v. 16, n. 10, p. 461–468.
- HAN, M.J. 2010. et al. Burst expansion, distribution, and diversification of MITEs in the silkworm genome. *BMC Genomics*, v. 11, n. 1.
- HAN, Y.; QIN, S.; WESSLER, S. R. 2013. Comparison of class 2 transposable elements at superfamily resolution reveals conserved and distinct features in cereal grass genomes. *BMC Genomics*, v. 14, n. 1.
- HANDLER, A. M.; GOMEZ, S. P. 1997. A new hobo, Ac, Tam3 transposable element, hopper, from *Bactrocera dorsalis* is distantly related to hobo and Ac. *Gene*, v. 185, n. 1, p. 133–135.
- HIKOSAKA, A.; KAWAHARA, A. 2004. Lineage-specific tandem repeats riding on a transposable element of MITE in *Xenopus* evolution: A new mechanism for creating simple sequence repeats. *J. Mol. Evol.* 59, 738–746.
- JACQUES, P.É.; JEYAKANI, J.; BOURQUE, G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genetics*, v. 9, n. 5, p. e1003504.
- JURKA, J. et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, v. 110, n. 1–4, p. 462–467.
- KAESSMANN, H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research*, v. 20, n. 10, p. 1313–1326.
- KING, M; ROFE, R. 1976. Karyotypic variation in the Australian gecko *Phyllodactylus marmoratus* (Gray) (Gekkonidae: Reptilia). *Chromosoma*, v.54, p.75–87.



LONG, M. et al. 2003. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics*, v. 4, n. 11, p. 865–875.

LYNCH, V. J. et al. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics*, v. 43, n. 11, p. 1154–1159.

MA, G. et al. 2022. Transposable elements in pluripotent stem cells and human disease. *Frontiers in Genetics*, v. 13.

MACFARLAN, T. S. et al. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, v. 487, n. 7405, p. 57–63.

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

NORONHA, R. C. R. et al. 2016. New insights of karyoevolution in the Amazonian turtles *Podocnemis expansa* and *Podocnemis unifilis* (Testudines, Podocnemidae). *Molecular Cytogenetics*, v. 9, n. 1.

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

PAÇO, A.; FREITAS, R.; VIEIRA-DA-SILVA, A. 2019. Conversion of DNA sequences: from a transposable element to a tandem repeat or to a gene. *Genes*, v. 10, n. 12.

PIÉGU, B. et al. 2015. A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution*. 86, 90–109.

PRIZON, A.C., et al. 2018. Chromosome spreading of the retrotransposable Rex-3 element and microsatellite repeats in karyotypes of the *Ancistrus* Populations. *Zebrafish*, v. 15, n. 5, p. 504–514.

QUESNEVILLE, H. 2020. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mobile DNA*, v. 11, n. 1.

ROGERS, R. L. et al. 2014. Genomic takeover by transposable elements in the Strawberry poison frog. *Molecular Biology and Evolution*.

RUBIN, E.; LITHWICK, G.; LEVY, A. A. 2001. Structure and evolution of the hAT transposon superfamily. *Genetics*, v. 158, n. 3, p. 949–957.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. 1989. Molecular cloning: a laboratory manual (No. Ed. 2). *Cold spring harbor laboratory press*.

SCALVENZI, T.; POLLET. 2014. N. Insights on genome size evolution from a miniature inverted repeat transposon driving a satellite DNA. *Mol. Phylogenet. Evol.* 81, 1–9.

SILVA, D. M. Z. A. et al. 2013. Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): dispersive location, association, and co-localization in the genome. *Genetica*, v. 141, n. 7-9, p. 329–336.

SINZELLE, L. et al. 2005. Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian *Xenopus tropicalis*. *Gene*, v. 349, p. 187–196.

SCHMID, M. 1978. Chromosome banding in Amphibia. I. Constitutive heterochromatin and nucleolus organizer regions in Bufo and Hyla. *Chromosoma*, v.66, p.361–368.

SUN, Y. et al. 2015. 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.

SUNDARAM, V. et al. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*, v. 24, n. 12, p. 1963–1976.

TEIXEIRA, W.G., et al. 2009. Organization of repeated dna elements in the genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. *Cytogenetic and Genome Research*.

THOMAS, J.; VADNAGARA, K.; PRITHAM, E. J. 2014. DINE-1, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). *Mobile DNA*, v. 5, n. 1.

TRALDI, J. B. et al. 2019. Chromosomal distribution of the retroelements Rex 1, Rex 3, and Rex 6 in species of the genus *Harttia* and *Hypostomus* (Siluriformes: Loricariidae). *Neotropical Ichthyology*, v. 17, n. 2.

TRAUT, W.; EICKHOF, U.; SCHORCH, J. C. 2001. Identification and analysis of sex chromosomes by comparative genomic hybridization (CGH). *Methods in Cell Science*, v. 23, p. 155–161.

TRIZZINO, M.; KAPUSTA, A.; BROWN, C. D. 2018. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics*, v. 19, n. 1.

TUNJIĆ-CVITANIĆ, M. et al. 2021. Satellitome analysis of the pacific oyster *Crassostrea gigas* reveals new pattern of satellite dna organization, highly scattered across the genome. *International Journal of Molecular Sciences*, v. 22, n. 13, p. 6798.

VALENTE, G.T.; et al. 2010. 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.

VENKATESH; NANDINI, B. 2020. Miniature inverted-repeat transposable elements (MITEs), derived insertional polymorphism as a tool of marker systems for molecular plant breeding. *Molecular Biology Reports*, v. 47, n. 4, p. 3155–3167.

VOLFF, J. N. et al. 1999. The non-LTR retrotransposon Rex3 from the fish *Xiphophorus* is widespread among teleosts. *Molecular Biology and Evolution*, v. 16, n. 11, p.1427– 1438.

VOLFF, J. N. et al. 2001. Non-LTR retrotransposons encoding a restriction enzymelike endonuclease in vertebrates. *Journal of Molecular Evolution*, v. 52, n. 4, p. 351–360.

WANG, J.; HUANG, J.; SHI, G. 2020. Retrotransposons in pluripotent stem cells. *Cell Regeneration*, v. 9, n. 1.

WICKER, T. et al. 2007. A unified classification system for eukaryotic transposable elements. *Nature genetics*, v. 8, p. 973–982.

YANG, H.-P.; BARBASH, D. A. 2008. Abundant and species-specific DINE-1 transposable elements in 12 *Drosophila* genomes. *Genome Biology*, v. 9, n. 2, p. R39.

YE, C.; JI, G.; LIANG, C. 2016. detectMITE: A novel approach to detect miniature inverted repeat transposable elements in genomes. *Scientific Reports*, v. 6, n. 1.

ZATTERA, M.L.; et al. 2020. Evolutionary Dynamics of the Repetitive DNA in the Karyotypes of *Pipa carvalhoi* and *Xenopus tropicalis* (Anura, Pipidae). *Front. Genet.* 11, 1–10.

ZATTERA, M. L.; BRUSCHI, D. P. 2022. Transposable elements as a source of novel repetitive DNA in the eukaryote genome. *Cells*, v. 11, n. 21, p. 3373.

ZELJKO, V. T. et al. 2020. Satellite DNA-like repeats are dispersed throughout the genome of the Pacific oyster *Crassostrea gigas* carried by Helentron non-autonomous mobile elements. *Scientific Reports*, v. 10, n. 1.

ZUO, B; et al. 2023. Comparative genomics reveals insights into anuran genome size evolution. *BMC Genomics*, vol. 24, no. 1.

### **3.3. CAPÍTULO 3:**

**DO THE INSERTION PREFERENCES OF TRANSPOSABLE ELEMENTS  
REPRESENT A STRATEGY FOR THE MITIGATION OF GENETIC  
CONFLICTS WITH THE HOST?**

## **Digest: Do the insertion preferences of transposable elements represent a strategy for the mitigation of genetic conflicts with the host?**

Michelle Louise Zattera<sup>1</sup> & Daniel Pacheco Bruschi<sup>1</sup>

<sup>1</sup> Universidade Federal do Paraná, Setor de Ciências Biológicas, Departamento de Genética, Programa de Pós-Graduação em Genética, Curitiba, PR, Brazil.

E-mails: [zatteralouise@gmail.com](mailto:zatteralouise@gmail.com); [danielpachecobruschi@gmail.com](mailto:danielpachecobruschi@gmail.com).

**Title of the original article:** Critical role of insertion preference for invasion trajectory of transposons.

**Abstract:** Transposable elements can exhibit a predilection for specific insertion regions. The article by Munasinghe et al. (2023) discusses insertions in positions considered "neutral" in the genome, emphasizing the relevance of the host restriction mechanisms, revealing that insertion in specific regions may be a strategy to mitigate genetic conflicts with the host.

### **3.3.1 Main text**

A key question in the evolution of the genome is why so many families of Transposable Elements (TEs), which have the potential to replicate rapidly and multiply in the host genome, tend to be found in relatively small or moderate numbers of copies in the genomes they have invaded. At a population level, Charlesworth (summarized in Charlesworth, 1994) proposed that the rate of transposition is counterbalanced by excision rates, together with selection against deleterious insertions or chromosomal rearrangements, which leads to an equilibrium in the number of copies found in the host genome. This equilibrium would be what allows the TEs to both persist in the host genome and spread throughout the reproductive population.

Over the three decades since Charlesworth's proposal, advances in the understanding of the biology of TEs, including the identification of preferential

insertion sites and intragenomic regulation mechanisms, have provided important insights into the mechanisms that determine the insertion and fixation of these elements in the genome. Using stochastic simulations, Munasinghe et al. (2023) verified how the preference for a specific genomic region can determine the impact the success of a novel TE invasion. Munasinghe et al. (2023) designed a diploid, hemaphroditic population that reproduces sexually composed by five chromosomes with uniform recombination rate to test the transposition impact of a unique TE family. The Authors designed evolutionary scenarios considered TEs with preference for either neutral or deleterious insertions combined with scenarios with or without host repression limiting the transposition rate.

The results of these simulations reinforced the conclusion that the success of the insertion depends on a certain level of cooperative interaction between the TE and the host genome. To begin with, the propensity of the TEs to insert themselves in regions that do not impact the host's adaptive potential would guarantee a balance between their potential for amplification and any deleterious effects arising from the presence of the TE in the genome. Even so, this evolutionary strategy may not be sufficient to guarantee the successful invasion of the host genome, according to the results of Munasinghe et al.'s (2023) simulation.

Even when the TEs are inserted in neutral regions, their transposition must be controlled subsequently silencing mechanisms to prevent them reaching the threshold at which deleterious effects may arise, leading to the extinction of both the host and the TEs. In this case, the response of the host to the transposition will be fundamental to the survival of both.

While the host will evolve to suppress the TEs, the latter will attempt to guarantee their cycle of replication during the invasion. However, this interaction may not represent an arms race in the sense of the "Red Queen" model (Van Valen, 1973), used traditionally to explain TE-host interactions, but rather by "good neighborhood" cooperative practices. In this case, the preference for neutral insertions would be an

evolutionary strategy adopted by some eukaryotic TE families to mitigate the genetic conflicts with their host. The initial replicative cycles in the euchromatin may thus represent an opportunity to maximize the number of copies generated before the host is able to recognize and silence the TEs (Zhang et al., 2020). This strategy would reduce the pressure on the host genome, with a temporary ceasefire in the arms race that allows for the generation of copies and guarantees that both the TE and the host survive and spread copies throughout the host population.

Overall, then, the results of Munasinghe et al. (2023) have provided important insights for the understanding of the processes that may have guided the evolution of insertion preferences as an integral component of the invasion and consolidation mechanisms of the TEs in the host population.

**Acknowledgements:** We would like to thank the Brazilian Coordination for Higher Education Personnel Training (CAPES/PROAP – Finance Code 001 and CAPES-Print) for the scholarships provided to MLZ. DPB is grateful to the Brazilian National Council for Scientific and Technological Development (CNPq #303646/2021-7, CNPq #401024/2022-9, and CNPq#405041/2021-7).

**Data Accessibility Statement:** There is no data archived.

## References

- Charlesworth, B., Sniegowski, P., Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*.
- Munasinghe, M., Springer, N., Brandvain, Y. 2023. Critical role of insertion preference for invasion trajectory of transposons. *Evolution*. doi: 10.1093/evolut/qpad128
- Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory*. doi: 10.7208/9780226115504-022
- Zhang, X., Zhao, M., McCarty, D.R., Lisch, D. Transposable elements employ distinct integration strategies with respect to transcriptional landscapes in eukaryotic genomes. *Nucleic Acids Research*. doi: 10.1093/nar/gkaa370.



#### 4. CONSIDERAÇÕES FINAIS

O presente estudo buscou elucidar as questões ao redor de elementos transponíveis como origem para sequências de DNA satélite dentro do genoma de anfíbios. No primeiro capítulo, foi realizada uma revisão de literatura aprofundada sobre casos previamente documentados de elementos transponíveis dando origens a outras sequências repetidas em tandem, levantando os principais mecanismos envolvidos nestes processos, bem como regiões do TE propensas para essa origem.

No segundo capítulo, o elemento transponível de hAT foi reportado pela primeira vez no genoma de *Pipa carvalhoi*, com mapeamento cromossômico e análises de bioinformática. Ainda, foi possível perceber a presença de repetições de satélite dentro deste elemento, fato também documentado pela primeira vez dentro da família Pipidae, nas espécies de *Pipa carvalhoi* e *Xenopus tropicalis*, levantando questionamentos acerca do papel destes elementos transponíveis no espalhamento de outras sequências repetitivas.

Em um terceiro momento, houve uma breve discussão à cerca da inserção preferencial dos elementos transponíveis em determinadas regiões, a fim de evitar conflitos com o hospedeiro onde este elemento se encontra inserido.

Os resultados obtidos até o presente momento contribuem para o entendimento da dinâmica genômica em anfíbios, ao passo que abrem caminho para investigações adicionais sobre o papel dos elementos transponíveis na geração de sequências de repetidas em tandem, como DNA satélites.

## REFERÊNCIAS

- AGUDO, M.; et al. Centromeres from telomeres? The centromeric region of the Y chromosome of *Drosophila melanogaster* contains a tandem array of telomeric HeT-A- and TART-related sequences. **Nucleic Acids Research**, v. 27, p. 3318–3324, 1999.
- AHMED, M.; LIANG, P. Transposable elements are a significant contributor to tandem repeats in the human genome. **Comparative and Functional Genomics**, 2012.
- ALI, A.; HAN, K.; LIANG, P. Role of transposable elements in gene regulation in the human genome. **Life**, v. 11, n. 2, p. 118, 2021.
- ALZOHAIRY, A.M. et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. **Plasmid**, v. 69, p. 1–15, 2013.
- AMORIM, I.C. et al. Comprehensive mapping of transposable elements reveals distinct patterns of element accumulation on chromosomes of wild beetles. **Chromosome Research**, v. 29, p. 203–218, 2021.
- ARCOT, S.S. et al. repeats: a source for the genesis of primate microsatellites. **Genomics**, v. 29, p. 136–144, 1995.
- ARMOUR, J.A.L. et al. Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat elements. **Nucleic Acids Research**, v. 17, p. 4925–4936, 1989.
- BALACHANDRAN, P. et al. Transposable element-mediated rearrangements are prevalent in human genomes. **Nature Communications**, v. 13, n. 1, p. 7115, 2022.
- BANKS, J.A.; MASSON, P.; FEDOROFF, N. Molecular mechanisms in the developmental regulation of the maize Suppressor-Mutator transposable element. **Genes & Development**, v. 2, p. 1364–1380, 1988.
- BAO, J.; YAN, W. Male germline control of transposable elements. **Biology of Reproduction**, v. 86, p. 162, 1–14, 2012.
- BATISTONI, R. et al. A tandemly repeated DNA family originated from SINE-related elements in the European plethodontid salamanders (Amphibia, Urodela). **Journal of Molecular Evolution**, v. 40, p. 608–615, 1995.
- BELYAYEV, A.; JOSEFIOVÁ, J.; JANDOVÁ, M. Transposons and satellite DNA: On the origin of the major satellite DNA family in the *Chenopodium* genome. **Mobile DNA**, v. 11, p. 20, 2020.
- BENNETZEN, J.L.; WANG, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. **Annual Review of Plant Biology**, v. 65, p. 505–530, 2014.
- BÖHNE, A. et al. Transposable elements as drivers of genomic and biological diversity in vertebrates. **Chromosome Research**, v. 16, n. 1, p. 203–215, 2008.
- BOIS, P.; WILLIAMSON, J.; BROWN, J.; DUBROVA, Y.E.; JEFFREYS, A.J. A novel unstable mouse VNTR family expanded from SINE B1 elements. **Genomics**, v. 49, p. 122–128, 1998.

- BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.
- BROOKFIELD, J.F.Y. Evolutionary forces generating sequence homogeneity and heterogeneity within retrotransposon families. **Cytogenetic and Genome Research**, v. 110, n. 1-4, p. 383–391, 2005.
- BRUSCHI, D.P.; BUSIN, C.S.; SIQUEIRA, S.; RECCO-PIMENTEL, S.M. Cytogenetic analysis of two species in the *Phyllomedusa hypochondrialis* group (Anura, Hylidae). **Hereditas**, v. 149, p. 34–40, 2012.
- BUREAU, T.E.; WESSLER, S.R. Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, p. 1411–1415, 1994.
- BUSCHIAZZO, E.; GEMMELL, N. J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. **BioEssays: news and reviews in molecular, cellular and developmental biology**, v. 28, n. 10, 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**, 1988.
- CARDUCCI, F. et al. Rex retroelements and teleost genomes: An Overview. **International Journal of Molecular Sciences**, v. 19, p. 1–15, 2018.
- CASAVANT, N. C. et al. The end of the LINE?: lack of recent Ll activity in a group of south american rodents. **Genetics**, v. 154, n. 4, p. 1809–1817, 2000.
- CASTANERA, R. et al. Amplification dynamics of miniature inverted-repeat transposable elements and their impact on rice trait variability. **The Plant Journal**, v. 107, n. 1, p. 118–13, 2021.
- 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.
- CHANDLER, V.; RIVIN, C.; WALBOT, V. Stable non-mutator stocks of maize have sequences homologous to the Mu1 transposable element. **Genetics**, v. 114, p. 1007–1021, 1986.
- CHANG, C.-H. et al. Islands of Retroelements Are Major Components of Drosophila Centromeres. **PLoS Biology**, v. 17, e3000241, 2019.
- CHARLESWORTH, B., SNIEGOWSKI, P., STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, 1994.
- CHEN, J. et al. P-MITE: a database for plant miniature inverted-repeat transposable elements. **Nucleic Acids Research**, v. 42, n. D1, p. D1176–D1181, 2013.
- CHEN, J. et al. Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice. **Nature Communications**, v. 10, n. 1, 2019.

- CHEN, J.; LU, L.; ROBB, S.M.C. Genomic Diversity Generated by a Transposable Element Burst in a Rice Recombinant Inbred Population. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, p. 26288–26297, 2020.
- CHENG, Z.-J.; MURATA, M. A Centromeric Tandem Repeat Family Originating From a Part of Ty3/ Gypsy -Retroelement in Wheat and Its Relatives. **Genetics**, v. 164, p. 665–672, 2003.
- CHOMET, P.S.; WESSLER, S.; DELLAPORTA, S.L. Inactivation of the Maize Transposable Element Activator (Ac) Is Associated with Its DNA Modification. **EMBO Journal**, v. 6, p. 295–302, 1987.
- CLARK, R.M. et al. Expansion of GAA Triplet Repeats in the Human Genome: Unique Origin of the FRDA Mutation at the Center of an Alu. **Genomics**, v. 83, p. 373–383, 2004.
- COATES, B. S. et al. A Helitron-Like transposon superfamily from lepidoptera disrupts (GAAA)<sub>n</sub> microsatellites and is responsible for flanking sequence similarity within a microsatellite family. **Journal of Molecular Evolution**, v. 70, n. 3, p. 275–288, 2010.
- COATES, B.S. et al. A Novel Class of Miniature Inverted Repeat Transposable Elements (MITEs) That Contain Hitchhiking (GTCY)<sub>n</sub> Microsatellites: Lepidopteran Mobile Microsatellites. **Insect Molecular Biology**, v. 20, p. 15–27, 2011.
- COLLINGS, C.K.; ANDERSON, J.N. Links between DNA Methylation and Nucleosome Occupancy in the Human Genome. **Epigenetics & Chromatin**, v. 10, p. 18, 2017.
- COSTA, G. W. W. F. et al. The evolutionary dynamics of ribosomal genes, histone H3, and transposable Rex elements in the genome of atlantic snappers. **The Journal of Heredity**, v. 107, n. 2, p. 173–180, 2016.
- DASILVA, C.; HADJI, H.; OZOUF-COSTAZ, C. 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, p. 13636–13641, 2002.
- DAVIS, A. J.; CHEN, D. J. DNA double strand break repair via non-homologous end-joining. **Translational cancer research**, v. 2, n. 3, p. 130–143, 2013.
- DEININGER, P. L.; BATZER, M. A. Alu repeats and human disease. **Molecular genetics and metabolism**, v. 67, n. 3, p. 183–193, 1999.
- DEININGER, P. Alu elements: Know the SINEs. **Genome Biology**, v. 12, p. 236, 2011.
- DEPRÁ, M. et al. Mar, a MITE family of HAT transposons in *Drosophila*. **Mobile DNA**, v. 3, p. 13, 2012.
- DEWANNIEUX, M.; ESNAULT, C.; HEIDMANN, T. LINE-mediated retrotransposition of marked Alu sequences. **Nature Genetics**, v. 35, p. 41–48, 2003.
- DIAS, G.B.; HERINGER, P.; SVARTMAN, M. Helitrons shaping the genomic architecture of *Drosophila*: Enrichment of DINE-TR1 in  $\alpha$ - and  $\beta$ -heterochromatin, satellite DNA emergence, and piRNA expression. **Chromosome Research**, v. 23, p. 597–613, 2015.

- DIAS, G.B. et al. Tetris is a foldback transposon that provided the building blocks for an emerging satellite DNA of *Drosophila virilis*. **Genome Biology and Evolution**, v. 6, p. 1302–1313, 2014.
- DOVER, G. Molecular drive: A cohesive mode of species evolution. **Nature**, v. 299, p. 111–117, 1982.
- DUFFY, A.J.; COLTMAN, D.W.; WRIGHT, J.M. Microsatellites at a common site in the second ORF of L1 elements in mammalian genomes. **Mammalian Genome**, v. 7, p. 386–387, 1996.
- EBBS, M.L.; BENDER, J. Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. **Plant Cell**, v. 18, p. 1166–1176, 2006.
- EDWARDS, R.J et al. Draft genome assembly of the invasive cane toad, *Rhinella marina*. **Giga Science**, 2018.
- ELBARBARY, R.A.; LUCAS, B.A.; MAQUAT, L.E. Retrotransposons as regulators of gene expression. **Science**, v. 351, p. aac7247, 2016.
- ELDER JR, J. F.; TURNER, B. J. Concerted evolution of repetitive DNA sequences in eukaryotes. **The quarterly review of biology**, v. 70, n. 3, p. 297–320, 1995.
- ELLIOTT, T. A.; GREGORY, T. R. Do larger genomes contain more diverse transposable elements? **BMC evolutionary biology**, v. 15, n. 1, 2015.
- ENGELS, W.R. et al. High-frequency P element loss in *Drosophila* is homolog dependent. **Cell**, v. 62, p. 515–525, 1990.
- ERNST, C.; ODOM, D.T.; KUTTER, C. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. **Nature Communications**, v. 8, p. 1411, 2017.
- ETCHEGARAY, E.; NAVILLE, M.; VOLFF, J.-N.; HAFTEK-TERREAU, Z. Transposable element-derived sequences in vertebrate development. **Mobile DNA**, v. 12, p. 1, 2021.
- FERNANDES, T. L. et al. Carrying progeny on the back: reproduction in the Brazilian aquatic frog *Pipa carvalhoi*. **South American Journal of Herpetology**, v. 6, n. 3, p. 161–176, 2011.
- FERNÁNDEZ-MEDINA, R.D.; RIBEIRO, J.M.C.; CARARETO, C.M.A. Losing identity: structural diversity of transposable elements belonging to different classes in the genome of *Anopheles gambiae*. **BMC Genomics**, v. 13, p. 272, 2012.
- FERRIS, P. J. Characterization of a *Chlamydomonas* transposon, Gulliver, resembling those in higher plants. **Genetics**, v. 122, n. 2, p. 363–377, 1989.
- FESCHOTTE, C.; SWAMY, L.; WESSLER, S.R. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). **Genetics**, v. 163, p. 747–758, 2003.
- FESCHOTTE, C.; PRITHAM, E.J. DNA transposons and the evolution of eukaryotic genomes. **Annual Review of Genetics**, v. 41, p. 331–368, 2007.
- FESCHOTTE, C. Transposable elements and the evolution of regulatory networks. **Nature Reviews Genetics**, v. 9, p. 397–405, 2008.

- FESCHOTTE, C. et al. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. **Mobile DNA II**, 2002.
- FROST, D.R. 2023. Amphibian species of the world: an online reference. **American Museum of Natural History, New York, USA**. Electronic Database accessible at <http://research.amnh.org/herpetology/amphibia/index.html>. Version 6.0.
- FUEYO, R.; JUDD, J.; FESCHOTTE, C.; WYSOCKA, J. Roles of transposable elements in the regulation of mammalian transcription. **Nature Reviews Molecular Cell Biology**, v. 23, p. 481–497, 2022.
- GAMA, J. M. et al. A genomic survey of LINE elements in Pipidae aquatic frogs shed light on Rex-elements evolution in these genomes. **Molecular Phylogenetics and Evolution**, Volume 168, 2022.
- GAO, D. et al. Transposons play an important role in the evolution and diversification of centromeres among closely related species. **Frontiers in Plant Science**, v. 6, p. 216, 2015.
- GARRIDO-RAMOS, M. Satellite DNA: An evolving topic. **Genes**, v. 8, p. 230, 2017.
- GAZOLLA, C. B; et al. Evolutionary dynamics of DIRS-like and Ngaro-like retrotransposons in *Xenopus laevis* and *Xenopus tropicalis* genomes. **G3 Genes|Genomes|Genetics**, Volume 12, Issue 2, 2022.
- GENT, J. I.; DAWE, R. K. RNA as a structural and regulatory component of the centromere. **Annual Review of Genetics**, v. 46, n. 1, p. 443–453, 2012.
- GEORGE, J.A. et al. Genomic organization of the *Drosophila* telomere retrotransposable elements. **Genome Research**, v. 16, p. 1231–1240, 2006.
- GILBERT, D. M. et al. Precise repair of mPing excision sites is facilitated by target site duplication derived microhomology. **Mobile DNA**, v. 6, n. 1, 2015.
- GILL, R. A. et al. On the role of transposable elements in the regulation of gene expression and subgenomic interactions in crop genomes. **Critical Reviews in Plant Sciences**, v. 40, n. 2, p. 157–189, 2021.
- GONG, Z. et al. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. **Plant Cell**, v. 24, p. 3559–3574, 2012.
- GRAY, Y. H. M. It takes two transposons to tango. **Trends in Genetics**, v. 16, n. 10, p. 461–468, 2000.
- GRŽAN, T. et al. CenH3 distribution reveals extended centromeres in the model beetle *Tribolium castaneum*. **PLoS Genetics**, v. 16, e1009115, 2020.
- GU, W.; ZHANG, F.; LUPSKI, J. R. Mechanisms for human genomic rearrangements. **PathoGenetics**, v. 1, n. 1, p. 4, 2008.
- GUO, X. et al. De novo centromere formation and centromeric sequence expansion in wheat and its wide hybrids. **PLoS Genetics**, v. 12, e1005997, 2016.
- HAN, M.-J. et al. Burst Expansion, Distribution and Diversification of MITes in the *Silkworm* Genome. **BMC Genomics**, v. 11, p. 520, 2010.

HAN, Y.; QIN, S.; WESSLER, S. R. Comparison of class 2 transposable elements at superfamily resolution reveals conserved and distinct features in cereal grass genomes. **BMC Genomics**, v. 14, n. 1, 2013.

HANDLER, A. M.; GOMEZ, S. P. A new hobo, Ac, Tam3 transposable element, hopper, from *Bactrocera dorsalis* is distantly related to hobo and Ac. **Gene**, v. 185, n. 1, p. 133–135, 1997.

HARTL, D.L.; LOZOVSKAYA, E.R.; NURMINSKY, D.I. What Restricts the Activity of Mariner-like Transposable Elements. **Trends in Genetics**, v. 13, p. 197–201, 1997.

HAYWARD, A.; GILBERT, C. Transposable elements. **Current Biology**, v. 32, n. 17, p. R904–R909, 2022.

HEDGES, D. J.; DEININGER, P. L. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 616, n. 1-2, p. 46–59, 2007.

HEIKKINEN, E.; LAUNONEN, V.; MÜLLER, E.; BACHMANN, L. The PvB370 BamHI Satellite DNA Family of the *Drosophila Virilis* Group and Its Evolutionary Relation to Mobile Dispersed Genetic PDV Elements. **Journal of Molecular Evolution**, v. 41, p. 604–614, 1995.

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

HIKOSAKA, A.; KAWAHARA, A. Lineage-specific tandem repeats riding on a transposable element of MITE in *Xenopus* evolution: A new mechanism for creating simple sequence repeats. **Journal of Molecular Evolution**, v. 59, p. 738–746, 2004.

IRISSARRI, I. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family *Pipidae*. **BMC Evol. Biol**, v.11, p.114, 2011.

JACQUES, P.É.; JEYAKANI, J.; BOURQUE, G. The majority of primate-specific regulatory sequences are derived from transposable elements. **PLoS Genetics**, v. 9, n. 5, p. e1003504, 2013.

JANSSEN, A.; COLMENARES, S.U.; KARPEN, G.H. Heterochromatin: Guardian of the Genome. **Annual Review of Cell and Developmental Biology**, v. 34, p. 265–288, 2018.

JIANG, J.; BIRCHLER, J.A.; PARROTT, W.A.; KELLY DAWE, R. A Molecular View of Plant Centromeres. **Trends in Plant Science**, v. 8, p. 570–575, 2003.

JOHNSON, M. E. et al. A specific family of interspersed repeats (SINES) facilitates meiotic synapsis in mammals. **Molecular cytogenetics**, v. 6, n. 1, 2013.

JURKA, J. et al. Repbase Update, a database of eukaryotic repetitive elements. **Cytogenetic and Genome Research**, v. 110, n. 1–4, p. 462–467, 2005.

JURKA, J.; GENTLES, A.J. Origin and Diversification of Minisatellites Derived from Human Alu Sequences. **Gene**, v. 365, p. 21–26, 2006.

KAESSMANN, H. Origins, evolution, and phenotypic impact of new genes. **Genome Research**, v. 20, n. 10, p. 1313–1326, 2010.

- KANHAYUWA, L.; COUTTS, R.H.A. Short Interspersed Nuclear Element (SINE) Sequences in the Genome of the Human Pathogenic Fungus *Aspergillus Fumigatus* Af293. **PLoS ONE**, v. 11, p. e0163215, 2016.
- KAPITONOV, V. V.; JURKA, J. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. **Springer eBooks**, p. 27–37, 2000.
- KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature Reviews Genetics**, 9(5), 411–412, 2008.
- KAPITONOV, V.V.; HOLMQUIST, G.P.; JURKA, J. L1 Repeat Is a Basic Unit of Heterochromatin Satellites in Cetaceans. **Molecular Biology and Evolution**, v. 15, p. 611–612, 1998.
- KAPITONOV, V.V.; JURKA, J. Helitrons on a Roll: Eukaryotic Rolling-Circle Transposons. **Trends in Genetics**, v. 23, p. 521–529, 2007.
- KAPITONOV, V.V.; JURKA, J. Molecular Paleontology of Transposable Elements from *Arabidopsis Thaliana*. In: McDonald, J.F. (Ed.). **Transposable Elements and Genome Evolution**. Springer: Dordrecht, The Netherlands, pp. 27–37, 2000.
- KAPITONOV, V.V.; JURKA, J. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. **PLoS Biology**, v. 3, p. E181, 2005.
- KAZAZIAN, H. H., Jr et al. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. **Nature**, v. 332, n. 6160, p. 164–166, 1988.
- KAZAZIAN, H.H. Mobile Elements: Drivers of Genome Evolution. **Science**, v. 303, p. 1626–1632, 2004.
- KELKAR, Y.D. et al. A Matter of Life or Death: How Microsatellites Emerge in and Vanish from the Human Genome. **Genome Research**, v. 21, p. 2038–2048, 2011.
- KELLY, R.G. Similar Origins of Two Mouse Minisatellites within Transposon-like LTRs. **Genomics**, v. 24, p. 509–515, 1994.
- KIDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes. **Genetica**, v. 115, n. 1, p. 49–63, 2002.
- KIDWELL, M.G.; LISCH, D.R. Perspective: Transposable elements, parasitic DNA, and genome evolution. **Evolution**, v. 55, p. 1–24, 2001.
- KIM, S. et al. Transposable element-mediated structural variation analysis in dog breeds using whole-genome sequencing. **Mammalian Genome**, v. 30, n. 9-10, p. 289–300, 2019.
- KING, M; ROFE, R. Karyotypic variation in the Australian gecko *Phyllodactylus marmoratus* (Gray) (Gekkonidae: Reptilia). **Chromosoma**, v.54, p.75–87, 1976.
- KLEIN, S. J.; O'NEILL, R. J. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. **Chromosome Research**, 2018.
- LANDER, E.S. et al. International Human Genome Sequencing Consortium. Initial Sequencing and Analysis of the Human Genome. **Nature**, v. 409, p. 860–921, 2001.
- LANGDON, T. et al. De Novo Evolution of Satellite DNA on the Rye B Chromosome. **Genetics**, v. 154, p. 869–884, 2000.



LAVIE, L.; et al. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. **Genome Research**, v. 14, p. 2253–2260, 2004.

LE ROUZIC, A.; CAPY, P. The First Steps of Transposable Elements Invasion. **Genetics**, v. 169, p. 1033–1043, 2005.

LEGENDRE, M. et al. Sequence-based estimation of minisatellite and microsatellite repeat variability. **Genome Research**, v. 17, n. 12, p. 1787–1796, 2007.

LEO, L.; MARCHETTI, M.; GIUNTA, S.; FANTI, L. Epigenetics as an Evolutionary Tool for Centromere Flexibility. **Genes**, v. 11, p. 809, 2020.

LI, J.; LEUNG, F.C. A CR1 Element Is Embedded in a Novel Tandem Repeat (*Hin FI Repeat*) within the Chicken Genome. **Genome**, v. 49, p. 97–103, 2006.

LILLO, F.; FARAONE, F. P.; VALVO, M. L. Can the introduction of *Xenopus laevis* affect native amphibian populations? Reduction of reproductive occurrence in presence of the invasive species. **Biological Invasions**, v. 13, n. 7, p. 1533–1541, 2011.

LIPPMAN, Z.; MAY, B.; YORDAN, C. Distinct Mechanisms Determine Transposon Inheritance and Methylation via Small Interfering RNA and Histone Modification. **PLoS Biology**, v. 1, p. E67, 2003.

LIU, J.; ALI, M.; ZHOU, Q. Establishment and Evolution of Heterochromatin. **Annals of the New York Academy of Sciences**, v. 1476, p. 59–77, 2020.

LONG, M. et al. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics*, v. 4, n. 11, p. 865–875, 2003.

LÓPEZ-FLORES, I. et al. The Molecular Phylogeny of Oysters Based on a Satellite DNA Related to Transposons. **Gene**, v. 339, p. 181–188, 2004.

LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M.A. The Repetitive DNA Content of Eukaryotic Genomes. In: GARRIDO-RAMOS, M.A. (Ed.). **Genome Dynamics**, v. 7, p. 1–28, 2012.

LU, C. et al. Miniature Inverted-Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in *Oryza Sativa*. **Molecular Biology and Evolution**, v. 29, p. 1005–1017, 2012.

LYNCH, V. J. et al. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. **Nature Genetics**, v. 43, n. 11, p. 1154–1159, 2011.

MA, G. et al. Transposable elements in pluripotent stem cells and human disease. **Frontiers in Genetics**, v. 13, 2022.

MACAS, J. et al. Hypervariable 3' UTR Region of Plant LTR-Retrotransposons as a Source of Novel Satellite Repeats. **Gene**, v. 448, p. 198–206, 2009.

MACFARLAN, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. **Nature**, v. 487, n. 7405, p. 57–63, 2012.

- MAHESHWARI, S. et al. Location in Arabidopsis Is Unaltered by Extreme Divergence in CENH3 Protein Sequence. **Genome Research**, v. 27, p. 471–478, 2017.
- MARSANO, R.M.; DIMITRI, P. Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements. **Cells**, v. 11, p. 761, 2022.
- MAUPETIT-MEHOUS, S.; VAURY, C. Transposon Reactivation in the Germline May Be Useful for Both Transposons and Their Host Genomes. **Cells**, v. 9, p. 1172, 2020.
- METCALFE, C. J.; CASANE, D. Accommodating the load: The transposable element content of very large genomes. **Mobile genetic elements**, v. 3, n. 2, p. e24775, 2013.
- MCCLINTOCK, B. The origin and behavior of mutable loci in maize. **Proceedings of the National Academy of Sciences of the United States of America**, v. 36, n. 6, p. 344–355, 1950.
- NAMBIAR, M.; CHUANG, Y.-C.; SMITH, G.R. Distributing Meiotic Crossovers for Optimal Fertility and Evolution. **DNA Repair**, v. 81, p. 102648, 2019.
- NASCIMENTO, J.; BALDO, D.; LOURENÇO, L. B. First insights on the retroelement *Rex1* in the cytogenetics of frogs. **Molecular Cytogenetics**, v. 8, n. 1, p. 1–9, 2015.
- NITTA, N.; FARMAN, M.L.; LEONG, S.A. Genome Organization of *Magnaporthe Grisea*: Integration of Genetic Maps, Clustering of Transposable Elements and Identification of Genome Duplications and Rearrangements. **Theoretical and Applied Genetics**, v. 95, p. 20–32, 1997.
- NORONHA, R. C. R. et al. New insights of karyoevolution in the Amazonian turtles *Podocnemis expansa* and *Podocnemis unifilis* (Testudines, Podocnemidae). **Molecular Cytogenetics**, v. 9, n. 1, 2016.
- OLIVEIRA, S.G.; BAO, W.; MARTINS, C. Horizontal Transfers of Mariner Transposons between Mammals, and Insects. **Mobile DNA**, v. 3, p. 14, 2012.
- OZOUF-COSTAZ, C. et al. Genome dynamics and chromosomal localization of the non-LTR retrotransposons *Rex1* and *Rex3* in Antarctic fish. v. 16, n. 1, p. 51–57, 2004.
- PACE, J.K.; FESCHOTTE, C. The Evolutionary History of Human DNA Transposons: Evidence for Intense Activity in the Primate Lineage. **Genome Research**, v. 17, p. 422–432, 2007.
- PAÇO, A. et al. The Puzzling Character of Repetitive DNA in *Phodopus* Genomes (Cricetidae, Rodentia). **Chromosome Research**, v. 23, p. 427–440, 2015.
- PAÇO, A.; FREITAS, R.; VIEIRA-DA-SILVA, A. Conversion of DNA Sequences: From a Transposable Element to a Tandem Repeat or to a Gene. **Genes**, v. 10, p. 1014, 2019.
- PALOMEQUE, T. et al. Detection of a Mariner-like Element and a Miniature Inverted-Repeat Transposable Element (MITE) Associated with the Heterochromatin from Ants of the Genus *Messor* and Their Possible Involvement for Satellite DNA Evolution. **Gene**, v. 371, p. 194–205, 2006.
- PASERO, P. et al. Long-Range Organization and Sequence-Directed Curvature of *Xenopus Laevis* Satellite 1 DNA. **Nucleic Acids Research**, v. 21, p. 4703–4710, 1993.
- PEREA-RESA, C.; BLOWER, M.D. Centromere Biology: Transcription Goes on Stage. **Molecular and Cellular Biology**, v. 38, e00263-18, 2018.

- PIÉGU, B et al. A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. **Molecular Phylogenetics and Evolution**, v. 86, p. 90–109, 2015.
- PLOHL, M. et al. Long-Term Conservation vs High Sequence Divergence: The Case of an Extraordinarily Old Satellite DNA in Bivalve Mollusks. **Heredity**, v. 104, p. 543–551, 2010.
- POUGH, F.H.; ANDREWS, R.M.; CADLE, J.E et al. **Herpetology, Second Edition**. New York: Prentice Hall, 2001.
- PRESTING, G.G. Centromeric Retrotransposons and Centromere Function. **Current Opinion in Genetics & Development**, v. 49, p. 79–84, 2018.
- PRITHAM, E.J.; FESCHOTTE, C. Massive Amplification of Rolling-Circle Transposons in the Lineage of the Bat *Myotis Lucifugus*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, p. 1895–1900, 2007.
- PRIZON, A.C., et al. Chromosome spreading of the retrotransposable *Rex-3* element and microsatellite repeats in karyotypes of the *Ancistrus* Populations. **Zebrafish**, v. 15, n. 5, p. 504–514, 2018.
- QIU, Y.; KÖHLER, C. Mobility Connects: Transposable Elements Wire New Transcriptional Networks by Transferring Transcription Factor Binding Motifs. **Biochemical Society Transactions**, v. 48, p. 1005–1017, 2020.
- QUESNEVILLE, H. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. **Mobile DNA**, v. 11, n. 1, 2020.
- RAMSAY, L. et al. Intimate Association of Microsatellite Repeats with Retrotransposons and Other Dispersed Repetitive Elements in *Barley (Hordeum)*. **The Plant Journal**, v. 17, p. 415–425, 1999.
- RECH, G. E. et al. Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. **Nature Communications**, v. 13, n. 1, p. 1948, 2022.
- RIGAL, M.; MATHIEU, O. A “Mille-Feuille” of Silencing: Epigenetic Control of Transposable Elements. **Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms**, v. 1809, p. 452–458, 2011.
- ROBBERECHT, C. et al. Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations. **Genome Research**, v. 23, n. 3, p. 411–418, 2012.
- ROGERS, R. L. et al. Genomic takeover by transposable elements in the *Strawberry poison frog (Oophaga)*. **Molecular Biology and Evolution**, 2014.
- ROSE, N.R.; KLOSE, R.J. Understanding the Relationship between DNA Methylation and Histone Lysine Methylation. **Biochimica et Biophysica Acta Gene Regulatory Mechanisms**, v. 1839, p. 1362–1372, 2014.

- ROSSI, M.S. et al. Retroviral-like Features in the Monomer of the Major Satellite DNA from the South American Rodents of the Genus *Ctenomys*. **DNA Sequence**, v. 3, p. 379–381, 1993.
- RUBIN, E.; LITHWICK, G.; LEVY, A. A. Structure and evolution of the hAT transposon superfamily. **Genetics**, v. 158, n. 3, p. 949–957, 2001.
- SAINT-LEANDRE, B.; NGUYEN, S.C.; LEVINE, M.T. Diversification and Collapse of a Telomere Elongation Mechanism. **Genome Research**, v. 29, p. 920–931, 2019.
- SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. Molecular cloning: a laboratory manual (No. Ed. 2). **Cold spring harbor laboratory press**, 1989.
- SATOVIĆ, E. et al. Adjacent Sequences Disclose Potential for Intra-Genomic Dispersal of Satellite DNA Repeats and Suggest a Complex Network with Transposable Elements. **BMC Genomics**, v. 17, p. 997, 2016.
- SCALVENZI, T.; POLLET, N. Insights on Genome Size Evolution from a Miniature Inverted Repeat Transposon Driving a Satellite DNA. **Molecular Phylogenetics and Evolution**, v. 81, p. 1–9, 2014.
- SCHMID, M. Chromosome banding in Amphibia. I. Constitutive heterochromatin and nucleolus organizer regions in *Bufo* and *Hyla*. *Chromosoma*, v.66, p.361–368, 1978.
- SHANG, W.-H. et al. Chromosome Engineering Allows the Efficient Isolation of Vertebrate Neocentromeres. **Developmental Cell**, v. 24, p. 635–648, 2013.
- SHANG, W.-H. et al. Chickens Possess Centromeres with Both Extended Tandem Repeats and Short Non-Tandem-Repetitive Sequences. **Genome Research**, v. 20, p. 1219–1228, 2010.
- SHARMA, A.; PRESTING, G.G. Evolution of Centromeric Retrotransposons in Grasses. **Genome Biology and Evolution**, v. 6, p. 1335–1352, 2014.
- SHARMA, A.; WOLFGRUBER, T.K.; PRESTING, G.G. Tandem Repeats Derived from Centromeric Retrotransposons. **BMC Genomics**, v. 14, p. 142, 2013.
- SHI, X.; SELUANOV, A.; GORBUNOVA, V. Cell Divisions Are Required for L1 Retrotransposition. **Molecular and Cellular Biology**, v. 27, p. 1264–1270, 2007.
- SIGMAN, M.J.; SLOTKIN, R.K. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. **Plant Cell**, v. 28, p. 304–313, 2016.
- SILVA, D. M. Z. A. et al. Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): dispersive location, association, and co-localization in the genome. **Genetica**, v. 141, n. 7-9, p. 329–336, 2013.
- SINZELLE, L. et al. Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian *Xenopus tropicalis*. **Gene**, v. 349, p. 187–196, 2005.
- SIRIJOVSKI, N. et al. NfCR1, the First Non-LTR Retrotransposon Characterized in the Australian Lungfish Genome, *Neoceratodus Forsteri*, Shows Similarities to CR1-like Elements. **Journal of Experimental Zoology Part B: Molecular and Developmental Evolution**, v. 304, p. 40–49, 2005.

- SLOTKIN, R.K.; MARTIENSSSEN, R. Transposable Elements and the Epigenetic Regulation of the Genome. **Nature Reviews Genetics**, v. 8, p. 272–285, 2007.
- SUNTRONPONG, A. et al. Characterization of Centromeric Satellite DNAs (MALREP) in the Asian Swamp Eel (*Monopterus Albus*) Suggests the Possible Origin of Repeats from Transposable Elements. **Genomics**, v. 112, p. 3097–3107, 2020.
- TALBERT, P.B.; HENIKOFF, S. What Makes a Centromere? **Experimental Cell Research**, v. 389, p. 111895, 2020.
- TEIXEIRA, W.G., et al. Organization of repeated DNA elements in the genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. **Cytogenetic and Genome Research**, 2009.
- TEK, A.L.; SONG, J.; MACAS, J. Sobo, a Recently Amplified Satellite Repeat of Potato, and Its Implications for the Origin of Tandemly Repeated Sequences. **Genetics**, v. 170, p. 1231–1238, 2005.
- THAKUR, J.; PACKIARAJ, J.; HENIKOFF, S. Sequence, Chromatin and Evolution of Satellite DNA. **International Journal of Molecular Sciences**, v. 22, p. 4309, 2021.
- THOMAS, J.; VADNAGARA, K.; PRITHAM, E. J. DINE-1, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). **Mobile DNA**, v. 5, n. 1, 2014.
- THON, M.R.; MARTIN, S.L.; GOFF, S. BAC End Sequences and a Physical Map Reveal Transposable Element Content and Clustering Patterns in the Genome of *Magnaporthe Grisea*. **Fungal Genetics and Biology**, v. 41, p. 657–666, 2004.
- TRALDI, J. B. et al. Chromosomal distribution of the retroelements Rex 1, Rex 3, and Rex 6 in species of the genus *Harttia* and *Hypostomus* (Siluriformes: Loricariidae). **Neotropical Ichthyology**, v. 17, n. 2, 2019.
- TRAUT, W.; EICKHOF, U.; SCHORCH, J. C. Identification and analysis of sex chromosomes by comparative genomic hybridization (CGH). **Methods in Cell Science**, v. 23, p. 155–161, 2001.
- TRIZZINO, M.; KAPUSTA, A.; BROWN, C. D. Transposable elements generate regulatory novelty in a tissue-specific fashion. **BMC Genomics**, v. 19, n. 1, 2018.
- TRUEB, L, ROSS, C.F, SMITH, R. A new pipoid anuran from the late cretaceous of South Africa. **Journal of Vertebrate Paleontology**, v.25, p.533–547, 2005.
- TUNJIĆ-CVITANIĆ, M. et al. Satellitome analysis of the pacific oyster *Crassostrea gigas* reveals new pattern of satellite DNA organization, highly scattered across the genome. **International Journal of Molecular Sciences**, v. 22, n. 13, p. 6798, 2021.
- TYMOWSKA, J.; FISCHBERG, M. Chromosome complements of the genus *Xenopus*. **Chromosoma**, v. 44, n. 3, p. 335–342, 1973.
- UNO, Y. et al. Homoeologous chromosomes of *Xenopus laevis* are highly conserved after whole-genome duplication. **Heredity**, v. 111, n. 5, p. 430–436, 2013.

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, 2010.

VAN VALEN, L. A new evolutionary law. **Evolutionary Theory**, 1973.

VENKATESH; NANDINI, B. Miniature inverted-repeat transposable elements (MITEs), derived insertional polymorphism as a tool of marker systems for molecular plant breeding. **Molecular Biology Reports**, v. 47, n. 4, p. 3155–3167, 2020,

VOLFF, J. N. 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.

VONDRÁK, T. et al. Characterization of Repeat Arrays in Ultra-Long Nanopore Reads Reveals Frequent Origin of Satellite DNA from Retrotransposon-Derived Tandem Repeats. **The Plant Journal**, v. 101, p. 484–500, 2020.

WALTER, M. et al. An Epigenetic Switch Ensures Transposon Repression upon Dynamic Loss of DNA Methylation in Embryonic Stem Cells. **eLife**, v. 5, p. e11418, 2016.

WANG, J.; HUANG, J.; SHI, G. Retrotransposons in pluripotent stem cells. **Cell Regeneration**, v. 9, n. 1, 2020.

WANG, K.; et al. African Lungfish Genome Sheds Light on the Vertebrate Water-to-Land Transition. **Cell**, v. 184, p. 1362–1376.e18, 2021.

WANG, L.; et al. Comparative Analysis of Transposable Elements and the Identification of Candidate Centromeric Elements in the Prunus Subgenus Cerasus and Its Relatives. **Genes**, v. 13, p. 641, 2022.

WARD, M.C.; ZHAO, S.; LUO, K. Silencing of Transposable Elements May Not Be a Major Driver of Regulatory Evolution in Primate iPSCs. **eLife**, v. 7, p. e33084, 2018.

WAWRZYNIAK, P.; PŁUCIENNICZAK, G.; BARTOSIK, D. The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. **Frontiers in Microbiology**, v. 8, p. 2353, 2017.

WICKBOM, T. The chromosomes of *Pipa pipa*. **Hereditas**, n.36, p.363-370, 1950.

WICKER, T.; SABOT, F.; HUA-VAN, A. A Unified Classification System for Eukaryotic Transposable Elements. **Nature Reviews Genetics**, v. 8, p. 973–982, 2007.

WOLFGRUBER, T.K. et al. Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. **PLoS Genetics**, v. 5, e1000743, 2009.

WONG, L.H.; CHOO, K.H.A. Evolutionary Dynamics of Transposable Elements at the Centromere. **Trends in Genetics**, v. 20, p. 611–616, 2004.

XIN, Y. et al. Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (*Morus notabilis*). **Mobile DNA**, v. 10, n. 1, 2019.

- YANG, F.; XI, R. Silencing Transposable Elements in the *Drosophila* Germline. **Cellular and Molecular Life Sciences**, v. 74, p. 435–448, 2017.
- YANG, H.-P.; BARBASH, D. A. Abundant and species-specific DINE-1 transposable elements in 12 *Drosophila* genomes. **Genome Biology**, v. 9, n. 2, p. R39, 2008.
- YE, C.; JI, G.; LIANG, C. detectMITE: A novel approach to detect miniature inverted repeat transposable elements in genomes. **Scientific reports**, v. 6, n. 1, p. 1–11, 2016.
- ZAMUDIO, N.; BOURC'HIS, D. Transposable Elements in the Mammalian Germline: A Comfortable Niche or a Deadly Trap? **Heredity**, v. 105, p. 92–104, 2010.
- ZATTERA, M. L. et al. Chromosome spreading of the (TTAGGG)<sub>n</sub> repeats in the *Pipa carvalhoi* Miranda-Ribeiro, 1937 (Pipidae, Anura) karyotype. **Comparative cytogenetics**, v. 13, n. 3, p. 297–309, 2019.
- ZATTERA, M. L. et al. Evolutionary Dynamics of the Repetitive DNA in the Karyotypes of *Pipa carvalhoi* and *Xenopus tropicalis* (Anura, Pipidae). **Frontiers in genetics**, v. 11, 2020.
- ZATTERA, M. L.; BRUSCHI, D. P. Transposable elements as a source of novel repetitive DNA in the eukaryote genome. **Cells**. v. 11, n. 21, p. 3373, 2022.
- ZELJKO, V. T. et al. 2020. Satellite DNA-like repeats are dispersed throughout the genome of the Pacific oyster *Crassostrea gigas* carried by Helentron non-autonomous mobile elements. **Scientific Reports**, v. 10, n. 1.
- ZELJKO, V.T.; PAVLEK, M.; MEŠTROVIĆ, N. Satellite DNA-like Repeats Are Dispersed throughout the Genome of the Pacific Oyster *Crassostrea Gigas* Carried by Helentron Non-Autonomous Mobile Elements. **Scientific Reports**, v. 10, p. 15107, 2020.
- ZHANG, H.H.; et al. A Novel HAT Element in *Bombyx Mori* and *Rhodnius Prolixus*: Its Relationship with Miniature Inverted Repeat Transposable Elements (MITEs) and Horizontal Transfer. **Insect Molecular Biology**, v. 22, p. 584–596, 2013.
- ZHANG, X. et al. Transposable elements employ distinct integration strategies with respect to transcriptional landscapes in eukaryotic genomes. **Nucleic acids research**, v. 48, n. 12, p. 6685–6698, 2020.
- ZHAO, Z. et al. Genome-wide analysis of Tandem repeats in plants and green algae. **G3 (Bethesda, Md.)**, v. 4, n. 1, p. 67–78, 2014.
- ZHU, L. et al. Short Tandem Repeats in plants: Genomic distribution and function prediction. **Electronic journal of biotechnology: EJB**, v. 50, p. 37–44, 2021.
- ZUO, B.; NNEJI, L. M.; SUN, Y.-B. Comparative genomics reveals insights into anuran genome size evolution. **BMC genomics**, v. 24, n. 1, 2023..