

UNIVERSIDADE FEDERAL DO PARANÁ

FERNANDA SOUZA DE OLIVEIRA

CITOGENÔMICA E CARACTERIZAÇÃO MOLECULAR DO TRANSPOSON DE DNA  
*hAT* EM *Apareiodon* sp. (CHARACIFORMES: PARODONTIDAE)

CURITIBA

2021

FERNANDA SOUZA DE OLIVEIRA

CITOGENÔMICA E CARACTERIZAÇÃO MOLECULAR DO TRANSPOSON DE DNA  
*hAT* EM *Apareiodon* sp. (CHARACIFORMES: PARODONTIDAE)

Dissertação apresentada ao curso de Pós-Graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Genética.

Orientador: Prof. Dr. Marcelo Ricardo Vicari

CURITIBA

2021

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

Oliveira, Fernanda Souza de.

Citogenômica e caracterização molecular do transposon de Dna *hAT* em *Apareiodon* sp. (Characiformes: Parodontidae). / Fernanda Souza de Oliveira. – Curitiba, 2021.

66 f. : il.

Orientador: Marcelo Ricardo Vicari.

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

1. Cromossomos sexuais. 2. Genomas. 3. Peixes tropicais. 4. Citogenética animal. I. Título. II. Vicari, Marcelo Ricardo. III. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 597



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

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **FERNANDA SOUZA DE OLIVEIRA** intitulada: **Citogenômica e caracterização molecular do transposon de DNA hAT em Apareiodon sp. (Characiformes: Parodontidae)**, sob orientação do Prof. Dr. MARCELO RICARDO VICARI, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

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

CURITIBA, 29 de Março de 2021.

Assinatura Eletrônica

29/03/2021 19:39:32.0

MARCELO RICARDO VICARI

Presidente da Banca Examinadora (UNIVERSIDADE ESTADUAL DE PONTA GROSSA)

Assinatura Eletrônica

30/03/2021 16:17:18.0

ADRIANA LUDWIG

Avaliador Externo (INSTITUTO CARLOS CHAGAS)

Assinatura Eletrônica

31/03/2021 21:12:50.0

CESAR MARTINS

Avaliador Externo (UNIVERSIDADE EST. PAULISTA JÚLIO DE MESQUITA FILHO)

Setor de Ciências Biológicas, Centro Politécnico - CURITIBA - Paraná - Brasil

CEP 81531-980 - Tel: (41) 3361-1587 - E-mail: ppg-gen@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 85864

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 85864

## AGRADECIMENTOS

Gostaria de agradecer a todo o apoio e incentivo recebidos no decorrer destes dois anos das pessoas e instituições envolvidas no desenvolvimento deste trabalho.

Ao Programa de Pós-Graduação em Genética da Universidade Federal do Paraná (UFPR) e a todo seu corpo docente pelo apoio, ensinamentos e condições necessárias para o desenvolvimento desta pesquisa.

Ao prof. Dr. Marcelo Ricardo Vicari pela sua orientação, apoio e incentivo, e pelos ensinamentos teóricos e práticos que foram fundamentais para a elaboração deste trabalho.

À Michelle Schemberger, Jordana Oliveira, Matheus Azambuja e Viviane Nascimento por toda a grande ajuda prestada com as análises de bioinformática que foram fundamentais para a realização deste trabalho.

Ao laboratório CBSF LAB da Universidade Estadual de Ponta Grossa (UEPG) por oferecer a infraestrutura e o suporte necessários para a realização deste trabalho.

Aos colegas do CBSF LAB pela ótima convivência e pelo apoio e disposição para tirar dúvidas nos procedimentos práticos do laboratório.

Aos meus pais Márcia Michaloski Souza e João Carlos Rufino de Oliveira por sempre apoiarem e incentivarem meus estudos.

Aos meus familiares e amigos, especialmente minha amiga Letícia, por todo o apoio e pelos ótimos momentos juntos.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Processo: 2015/16661-1) e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Processo: 305142/2019-4) pelo aporte financeiro que ajudou a custear a pesquisa.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), pelo auxílio financeiro com a concessão da bolsa.

## RESUMO

Estudos envolvendo elementos transponíveis (TEs) têm demonstrado o importante papel destas sequências para a evolução e diversificação dos genomas eucarióticos. Peixes geralmente apresentam grande abundância e diversidade de transposons de DNA, os quais foram descritos a influenciar na evolução dos genomas. Um estudo genômico na espécie *Apareiodon* sp. demonstrou a presença de diversos grupos de TEs, sendo os transposons de DNA os mais abundantes, entre esses elementos pertencentes à superfamília *hAT*. Assim, o objetivo deste trabalho foi realizar uma caracterização genômica da abundância, degeneração molecular e da presença de diferentes famílias de elementos *hAT* em *Apareiodon* sp. e, associar estes dados às modificações genômicas e localização cariotípica. Para alcançar esse objetivo foi realizada a caracterização molecular de TEs *hAT* recuperados do genoma de *Apareiodon* sp., bem como análises de árvores de agrupamentos das sequências obtidas com elementos das diferentes famílias *hAT* descritas na literatura, estimativas de tempo de inserção e degeneração de sequências, avaliação da presença de miniaturas invertidas de elementos transponíveis (MITEs) e de sequências repetidas *in tandem* associadas a *hAT* e a localização *in situ* de sítios destes elementos nos cromossomos de *Apareiodon* sp. Os resultados obtidos revelaram que ~0,4% do genoma desta espécie corresponde a TEs da superfamília *hAT*, sendo a maioria de suas cópias não autônomas ou degeneradas, porém com a presença de cinco cópias que apresentaram todas as características necessárias para serem elementos autônomos. Análises de agrupamentos de elementos *hAT* demonstraram a presença de quatro grupos predominantes no genoma de *Apareiodon* sp. (Ac, Blackjack, Charlie e Tip100), sendo o grupo Ac o mais abundante e o grupo Charlie o que apresentou as sequências mais conservadas e possivelmente autônomas. A análise dos períodos de inserção destes TEs no genoma demonstra dois períodos principais de invasão de TEs *hAT*, um com menor invasão de sequências Ac, Charlie e Tip100 e, uma onda recente com grande disseminação de elementos Ac, Blackjack e Charlie. Cópias MITEs foram encontradas para estes quatro grupos, as quais apresentaram sequências TIRs bem conservadas entre si, sendo o grupo Blackjack o que apresentou o maior número de sequências. Em cópias degeneradas foi possível observar a presença de sequências microssatélite internas aos TEs, demonstrando uma provável geração de DNAs microssatélites a partir de *hAT*. Por fim, a análise da localização *in situ* destes TEs nos cromossomos de *Apareiodon* sp. demonstra uma distribuição dispersa destas sequências, com pequenos acúmulos em regiões terminais. A análise do conjunto dos dados obtidos demonstra que os TEs *hAT* de *Apareiodon* sp. estão no início da fase de senescência, com a grande maioria das cópias em diferentes níveis de degeneração, servindo de substrato para a geração de sequências MITEs ou de sequências microssatélites, porém ainda com a possibilidade de autonomia de elementos da família Charlie, onde foram observadas sequências possivelmente autônomas. No nível citogenético, embora o cromossomo sexual W de *Apareiodon* sp. ter se diferenciado por acúmulos de TEs a cerca de 20 milhões de anos, a invasão das sequências *hAT* ocorrida em torno de 5 milhões de anos atrás gerou apenas acúmulos terminais nos cromossomos Z e W, similar ao ocorrido nos autossomos. Por fim, a grande quantidade de sequências *hAT* presentes no genoma servirão para avaliar a possibilidade de eventos de coopção molecular para geração de possíveis sequências funcionais no genoma de *Apareiodon* sp.

Palavras-chave: Cromossomos sexuais, elementos transponíveis, evolução do genoma, peixes neotropicais, transposases.

## ABSTRACT

Studies concerning transposable elements (TEs) have demonstrated an important role of these sequences for the evolution and diversification of the eukaryote genomes. Genomic data in fishes usually show a great abundance of DNA transposons, which have been pointed to act in genomic novelties. A genomic study in *Apareiodon* sp. showed the occurrence of several TE groups, with the DNA transposons being the most abundant, and among these, elements belonging to the *hAT* superfamily. The main goal of this study was to perform a genomic count of abundance, molecular deterioration and the presence of different *hAT* families in *Apareiodon* sp., as well as, correlate this data to genomic changes and karyotype localization. To reach this goal the molecular characterization of *hAT* TEs recovered from the genome of *Apareiodon* sp. was carried out, as well as grouping analyze of *Apareiodon* sp. *hAT* sequences with those recognized in the literature, evaluation of insertion time and sequence degeneration, evaluation of the presence of miniature inverted repeat transposable elements (MITEs) and tandem repeats associated with *hAT*, and the *in situ* localization of these elements in *Apareiodon* sp. chromosomes. The obtained data revealed that ~0,4% of *Apareiodon* sp. genome corresponds to TEs of the *hAT* superfamily, being most of their copies non-autonomous and deteriorated, however with the presence of five copies that had all the characteristics necessary to be autonomous elements. Grouping analyzes of *hAT* elements demonstrated the presence of four predominant groups in *Apareiodon* sp. genome (Ac, Blackjack, Charlie and Tip100), where Ac elements were the most abundant and Charlie copies having the most conserved and possibly autonomous sequences. The genomic insertion time estimative of *hAT* elements in the genome showed two main periods of burst in the *Apareiodon* lineage. The first one was represented by few copies of Ac, Charlie and Tip100 elements and, the more recent wave showing a large dissemination of Ac, Blackjack and Charlie elements. MITEs copies similar to these four groups were detected in the genome possessing highly conserved TIRs, being the Blackjack group the most abundant. Microsatellite clusters were detected within deteriorated *hAT* elements suggesting a probable origin of microsatellites from the *hAT* sequences. The *in situ* localization experiments located dispersed *hAT* copies in the *Apareiodon* sp. chromosomes, besides some accumulation in terminal regions. The pool of data suggest that *hAT* sequences are beginning the senescence phase in *Apareiodon* sp. genome, with the vast majority of copies at different levels of degeneration, serving as a substrate for the generation of MITEs or microsatellite sequences, but still with the possibility of autonomy for elements of the Charlie family, where possibly autonomous sequences were observed. At cytogenetic level, while the W sex chromosome of *Apareiodon* sp. has been proposed to differentiate by TE accumulation around 20 million years ago, the burst of *hAT* sequences around 5 million years ago only generated dispersed terminal copies in Z and W chromosomes, similarly to the occurred in autosomes. Finally, the abundance of *hAT* sequences present in the genome could be useful to evaluate the possibility of molecular cooption events for the generation of possible functional sequences in *Apareiodon* sp. genome.

Keywords: Genome evolution, neotropical fishes, sex chromosomes, transposable elements, transposases.

## LISTA DE FIGURAS

- Figura 1 – Representação da região codificante da sequência de nucleotídeos da transposase de TEs *hAT* putativamente íntegros no genoma de *Apareiodon* sp. Indicados na legenda o sítio de início da transcrição (TSS), éxon, e sinal de poliadenilação (PoliA) ..... 35
- Figura 2 – Estrutura dos elementos *hAT* e sequências de aminoácidos da sua transposase. Em (a) estrutura geral dos TEs *hAT*, com os quatro domínios característicos de sua transposase (*BED*, *DNA binding*, *Catalytic* e *Insertion*) descritos na literatura flanqueados pelas sequências das TIRs. Os principais aminoácidos altamente conservados entre transposases *hAT*, de acordo com Atkinson (2014) e Yuan e Wessler (2011) apresentam-se destacados abaixo de cada domínio, e os resíduos catalíticos DDE apresentam-se destacados acima do domínio catalítico. Em (b) alinhamento de nove sequências conservadas da transposase de *hAT* de *Apareiodon* sp. pelo software MUSCLE. Os quatro domínios característicos da transposase estão indicados abaixo das sequências. Os resíduos de aminoácidos conservados entre todas as sequências estão indicados por (\*), e os resíduos altamente conservados entre as transposases de *hAT*, responsáveis por funções catalíticas ou de ligação às TIRs apresentam-se destacados em verde ..... 36
- Figura 3 – Alinhamento das cinco sequências mais conservadas das transposases *hAT* de *Apareiodon* sp., as quais apresentaram todos os resíduos de aminoácidos necessários para a atividade de transposição de acordo com Atkinson (2014) e Yuan e Wessler (2011) conservados. Os quatro domínios característicos da transposase estão indicados abaixo das sequências. Os resíduos de aminoácidos conservados entre todas as sequências estão indicados por (\*), e os resíduos altamente conservados entre as transposases de *hAT*, responsáveis por funções catalíticas ou de ligação às TIRs apresentam-se destacados em verde ..... 37
- Figura 4 – Agrupamentos de sequências utilizando o método Maximum Likelihood (ML) com as sequências de aminoácidos das transposases *hAT* de *Apareiodon* sp. e sequências de referência pertencentes às principais famílias identificadas para elementos *hAT*. Em a) agrupamento das nove sequências ditas conservadas e que apresentaram todos os domínios pertencentes a *hAT* em *Apareiodon* sp. Em b) agrupamento das sequências com a presença de domínios de dimerização C-terminais conservados de *Apareiodon* sp. Os quatro grupos principais obtidos (Ac, Blackjack, Charlie e Tip100) foram representados em cores nos ramos dos agrupamentos ..... 42
- Figura 5 – Estimativa de abundância e do tempo de inserção dos elementos correspondentes às famílias Ac, Blackjack, Charlie e Tip100 (superfamília *hAT*) em *Apareiodon* sp. Os picos indicam as ondas de invasão dos diferentes grupos de *hAT* nos genomas macho (a) e fêmea (b) de *Apareiodon* sp. O eixo X indica estimativa do período de invasão dos elementos nos genomas (em milhões de anos), enquanto o eixo Y indica a abundância (%) destas sequências no conjunto total de sequências *hAT* presentes no genoma de *Apareiodon* sp. .... 43
- Figura 6 – Distribuição de DNAs microssatélite mais abundantes (> 1% dentro da biblioteca da superfamília *hAT*) pelas sequências dos TEs *hAT* de *Apareiodon* sp. dos genomas macho (azul) e fêmea (vermelho) ..... 47
- Figura 7 – Cinco exemplos de contigs do genoma fêmea de *Apareiodon* sp. analisados através do software Dfam demonstrando a presença de sequências repetidas *in tandem* associadas a

sequências *hAT*. São destacadas com \* e círculos em vermelho as sequências específicas de TEs do tipo *hAT* (laranja) com a presença de associações com sequências repetidas *in tandem* (preto). As identificações dos contigs são indicadas acima de cada imagem..... 48

Figura 8 – Cariótipo de *Apareiodon* sp. submetido à técnica de FISH com sonda da superfamília *hAT* (vermelho). Os cromossomos foram contracolorados com DAPI. Barra = 10  $\mu$ m ..... 49

## LISTA DE TABELAS

Tabela 1: Sequências de referência obtidas através dos bancos de dados do Dfam e do NCBI utilizadas para os agrupamentos de sequências <i>hAT</i> de <i>Apareiodon</i> sp. Indicados os números de acesso e IDs das sequências utilizadas, suas respectivas espécies e os grupos aos quais cada sequência pertence.....	28
Tabela 2: Descrição dos primers, regiões alvo e tamanho de amplicons na amplificação de <i>hAT</i> no genoma de <i>Apareiodon</i> sp. ....	30
Tabela 3: Abundância dos domínios característicos de <i>hAT</i> presentes no banco de dados do Pfam nas sequências analisadas. Contagem dos três domínios característicos de <i>hAT</i> com integridade de suas sequências maior e menor do que 90% e e-value <0,01 no genoma fêmea .....	34
Tabela 4: Sequências TIRs e TSDs encontradas nos TEs <i>hAT</i> de <i>Apareiodon</i> sp. Sequências das TIRs 5' e 3' flanqueando as sequências das transposases putativamente íntegras de cinco sequências <i>hAT</i> e sequências TSD observadas adjacentes às TIRs. Destacadas em negrito as bases conservadas tanto na extremidade 5' quanto na extremidade 3' .....	39
Tabela 5: Sequências de TIRs e TSDs obtidas para os elementos MITEs pertencentes ao grupo dos TEs <i>hAT</i> de <i>Apareiodon</i> sp. As sequências foram agrupadas de acordo com as quatro subfamílias mais abundantes observadas (Ac, Blackjack, Charlie e Tip100).....	45
Tabela 6: Estimativa da abundância (% dentro das sequências pertencentes a superfamília <i>hAT</i> ) de sequências microssatélites encontradas nas bibliotecas <i>hAT</i> de macho e fêmea de <i>Apareiodon</i> sp.....	46

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>12</b>
1.1 Elementos transponíveis .....	12
1.2 Superfamília <i>hAT</i> .....	15
1.3 TEs na evolução do genoma .....	17
1.3.1 Atuação dos elementos transponíveis na evolução dos cromossomos sexuais .....	20
1.3.2 Estudos moleculares em <i>Apareiodon</i> sp .....	21
<b>2. JUSTIFICATIVA E OBJETIVOS .....</b>	<b>24</b>
<b>3. MATERIAL E MÉTODOS .....</b>	<b>26</b>
3.1 Seleção e caracterização das sequências .....	26
3.2 Construção do Repeat Landscape .....	27
3.3 Agrupamentos de similaridade de sequências .....	28
3.4 Análise de sequências MITes .....	29
3.5 Avaliação da presença de repetições <i>in tandem</i> associadas à <i>hAT</i> .....	29
3.6 Obtenção de cromossomos mitóticos .....	30
3.7 Obtenção das sequências e geração de sondas .....	30
3.8 Localização <i>in situ</i> .....	31
3.9 Organização dos cariótipos .....	32
<b>4. RESULTADOS .....</b>	<b>33</b>
4.1 Caracterização molecular do elemento <i>hAT</i> .....	33
4.2 Análise da presença de elementos <i>hAT</i> íntegros em <i>Apareiodon</i> sp .....	34
4.3 Agrupamentos de similaridade e identificação de diversidade de sequências .....	40
4.4 Análise da invasão do TE <i>hAT</i> no genoma .....	42
4.5 Análise de sequências MITes .....	44
4.6 Análise da presença de sequências repetidas <i>in tandem</i> associadas à <i>hAT</i> .....	45
4.7 Localização <i>in situ</i> .....	48
<b>5. DISCUSSÃO .....</b>	<b>50</b>
5.1 Caracterização molecular dos TEs <i>hAT</i> no genoma de <i>Apareiodon</i> sp .....	50
5.2 Agrupamentos e análise de diversidade e elementos <i>hAT</i> em <i>Apareiodon</i> sp .....	53
5.3 Estimativa do tempo de inserção dos elementos <i>hAT</i> .....	54
5.4 Análise das sequências MITes pertencentes ao grupo dos TEs <i>hAT</i> .....	55
5.5 Associação de sequências repetidas <i>in tandem</i> à elementos <i>hAT</i> degenerados .....	56
5.6 Localização <i>in situ</i> de TEs <i>hAT</i> nos cromossomos de <i>Apareiodon</i> sp .....	57
<b>6. CONCLUSÃO .....</b>	<b>59</b>

<b>REFERÊNCIAS .....</b>	<b>60</b>
--------------------------	-----------

## 1. INTRODUÇÃO

### 1.1 Elementos transponíveis

Os DNAs repetitivos compreendem uma grande porção dos genomas eucarióticos, tendo funções estruturais e organizacionais, e estando envolvidos em mecanismos que geram diferenciação e rearranjos cromossômicos, sendo portanto, importantes fatores para a evolução do genoma (BISCOTTI; OLMO; HESLOP-HARRISON, 2015). Esses DNAs repetitivos são considerados *hotspots* para as quebras e rearranjos cromossômicos devido à instabilidade presente nessas regiões do genoma (PEVZNER; TESLER, 2003). Portanto, o estudo destas sequências implica em um maior conhecimento sobre a estrutura e funções genômicas em eucariotos (ARKHIPOVA, 2017).

Os elementos transponíveis (TEs) são sequências repetitivas que compõem uma porção significativa do genoma de eucariotos (WICKER et al., 2007). Estão geralmente presentes como cópias não autônomas, geradas por um processo de substituições/inserções e deleções nucleotídicas conhecido como degeneração molecular (FERNÁNDEZ-MEDINA et al., 2012). Eles podem estar presentes em um número pequeno de cópias ou podem chegar a apresentar centenas ou até milhares de cópias, neste último caso podendo representar uma fração significativa do genoma (KIDWELL; LISCH, 2001). Sua habilidade de se mover pelo genoma e se replicar conferem a estes elementos a possibilidade de contribuírem assim para o tamanho, estrutura e diversidade do genoma hospedeiro (KIDWELL, 2002).

A abundância de TEs parece ser uma das maiores contribuições para o tamanho e diversidade dos genomas, principalmente em peixes, onde eles podem representar de 6 à 55% do conteúdo do genoma (CHALOPIN et al., 2015a). Geralmente as quantidades de TEs nos genomas de vertebrados estão relacionadas com a idade do grupo taxonômico, onde grupos que já existem a períodos mais longos, como peixes, apresentam maior diversidade de TEs em comparação com grupos mais recentes, como mamíferos e aves (SOTERO-CAIO et al., 2017). Entretanto, as diferenças de quantidades de TEs entre espécies filogeneticamente próximas são também significativas (FERNÁNDEZ-MEDINA et al., 2012).

Os TEs podem ser agrupados em duas classes de acordo com seu mecanismo de transposição, sendo os elementos de classe I compostos pelos retrotransposons e, os elementos de classe II compostos pelos transposons de DNA (FINNEGAN, 1989). Além destas divisões há também subdivisões de classificação destes elementos em subclasses, ordens, superfamílias, famílias e subfamílias (WICKER et al., 2007).

A Classe I compreende os elementos que se transpõem de uma região à outra no genoma através de um intermediário de RNA, em um mecanismo denominado de copia-e-cola (FINNEGAN, 1989). Essa classe agrupa os elementos com longas repetições terminais (LTR) e os não-LTRs, os quais não apresentam as longas repetições terminais (WICKER et al., 2007). Os LTRs geralmente codificam uma proteína gag (*capsid protein*) e as enzimas ap (*aspartic proteinase*), rt (*reverse transcriptase*), rh (RNase H) e int (integrase) (WICKER et al., 2007). Os elementos LTR movem-se no genoma pelo mecanismo de transcrição em mRNA, conversão em cDNA pela proteína rt, seguida da inserção em um novo loco com atuação da int (HAVECKER; GAO; VOYTAS, 2004).

Elementos não-LTR geralmente apresentam uma cauda poli (A) ou repetições simples na região terminal 3' (KOJIMA, 2018). Entre os não-LTR estão inclusos os LINEs (*long interspersed nuclear element*), os quais codificam as proteínas rt e uma nuclease, e suas versões não autônomas SINEs (*short interspersed nuclear element*), que não codificam proteínas e dependem de elementos do tipo LINE para se moverem pelo genoma (EICKBUSH; JAMBURUTHUGODA, 2008; WICKER et al., 2007). Esses TEs transpõe-se pelo mecanismo em que o RNA intermediário se ancora ao DNA-alvo antes de ser convertido em cDNA (OSTERTAG; KAZAZIAN, 2001) promovendo um mecanismo simultâneo de transcrição reversa e inserção da cópia em um novo loco. Além destes grupos, os elementos pertencentes às ordens DIRS (*tyrosine recombinase-encoding retrotransposons*) e PLE (*Penelope-like elements*) apresentam estruturas e formas de transposição um pouco diferentes (EICKBUSH; JAMBURUTHUGODA, 2008; WICKER et al., 2007).

Os elementos da Classe II são caracterizados por se moverem diretamente de DNA para DNA, e ainda são classificados em duas subclasses devido a diferenças no mecanismo de transposição (WICKER et al., 2007). Para a subclasse I a transposição ocorre por mecanismo de corta-e-cola, no qual ocorre a clivagem do elemento através de proteínas específicas que realizam o corte das duas fitas de DNA e posterior transferência para outra região do genoma. Diferentemente do que ocorre para os retrotransposons, os transposons de DNA que se movimentam por mecanismo de corta-e-cola se multiplicam no genoma de diferentes formas, dependendo de mecanismos do hospedeiro, sendo estes mecanismos: a transposição durante a duplicação do DNA de uma região já duplicada para uma que ainda não passou pelo evento levando à geração de uma nova cópia; ou ainda, se o TE apresentar outra cópia no cromossomo homólogo, o reparo de quebra de dupla fita deixado após excisão do elemento pode gerar uma

nova cópia do TE através de recombinação entre o local excisado com o elemento do cromossomo homólogo (FESCHOTTE; PRITHAM, 2007).

Os elementos da subclasse I estão divididos em duas ordens entre os que apresentam TIRs (*terminal inverted repeats*) e os elementos Crypton, que codificam uma enzima tirosina recombinase (WICKER et al., 2007). Os elementos da ordem TIR também se caracterizam por apresentarem um gene para a enzima transposase, responsável pelo processo de transposição destes elementos (KOJIMA, 2018). Estes elementos estão divididos em superfamílias baseado na sequência de suas TIRs, no tamanho das TSDs (*target site duplications*) geradas após sua inserção (WICKER et al., 2007), além da similaridade das sequências que codificam a enzima transposase (YUAN; WESSLER, 2011).

Os transposons de DNA pertencentes à subclasse II, correspondentes aos elementos *Helitron* e *Maverick*, transpõe-se através de um mecanismo de copia-e-cola, no qual pode haver a replicação de uma ou duas fitas de DNA para sua posterior transferência (WICKER et al., 2007). Para esses TEs, proteínas específicas realizam o corte de apenas uma fita do DNA, seguida pela transposição desta fita à outro local genômico onde ocorre a síntese da outra fita (WICKER et al., 2007). Entre estes elementos os *Helitrons* se caracterizam por apresentarem um domínio para proteínas Rep/helicase para sua transposição, enquanto os elementos *Maverick* apresentam um gene para DNA polimerase B (KOJIMA, 2018; WICKER et al., 2007).

Tanto os TEs de Classe I quanto os de Classe II ainda podem ser divididos em autônomos, quando apresentam todos os domínios necessários para sua transposição, ou não autônomos, quando dependem de produtos de outros TEs para se moverem (WICKER et al., 2007). Quase todas as famílias de TEs apresentam elementos não autônomos, os quais podem ser gerados pela substituição de bases ou deleção de sequências internas, ou ainda, mantendo apenas fragmentos curtos de DNA entre suas sequências terminais, as quais podem ser reconhecidas por enzimas de outros elementos semelhantes (KOJIMA, 2018). Entre os transposons de DNA da ordem das TIRs ocorrem estes elementos curtos e não autônomos, denominados Miniaturas de Elementos Transponíveis (MITEs – *Miniature Inverted-repeat Transposable Elements*), as quais são caracterizadas por um tamanho entre 50 e 800 pb, sendo compostas por sequências de TIRs e TSDs iguais ou muito similares às de seus elementos autônomos (FATTASH et al., 2013; GUO et al., 2017). Essas MITEs geralmente apresentam um alto número de cópias nos genomas eucarióticos e podem ser agrupadas nas superfamílias

conhecidas dos transposons de DNA com base nas características de suas TIRs e TSDs (FATTASH et al., 2013).

Em diferentes genomas a natureza dos TEs presentes pode variar consideravelmente, algumas classes e famílias podem estar mais presentes em algumas espécies, enquanto outras podem estar mais presentes em outras (BISCOTTI; OLMO; HESLOP-HARRISON, 2015). Em peixes, por exemplo, há uma maior abundância de transposons de DNA em boa parte dos genomas analisados, em contrapartida em mamíferos há maior abundância de retrotransposons (CHALOPIN et al., 2015b; FESCHOTTE; PRITHAM, 2007). Ainda foi observado que para peixes teleósteos o tamanho do genoma é proporcional à quantidade de TEs identificados em cada espécie, sendo os genomas maiores encontrados em espécies que tiveram maior invasão de TEs (GAO et al., 2016).

## 1.2 Superfamília *hAT*

A superfamília *hAT* pertence à classe II de transposons de DNA e compreende os primeiros TEs identificados, os elementos *Ac* (*Activator*) em milho (MCCLINTOCK, 1947), e foi nomeada com base nos três elementos primeiramente descritos (*hobo*, *Ac* e *Tam3*) (WARREN; ATKINSON; O'BROCHTA, 1994). Apesar dos transposons *hAT* serem mais comuns em plantas do que em animais eles estão amplamente distribuídos em todos os reinos em eucariotos, onde pesquisadores sugerem que estes elementos provavelmente já existiam no início do processo de separação de plantas, animais e fungos, e que então se espalharam por estes grupos (RUBIN; LITHWICK; LEVY, 2001).

Características comuns a todos os transposons desta superfamília são: a geração de uma TSD de 8 pb no local de sua inserção, a presença de TIRs curtas, e assim como para todos os transposons de DNA autônomos, a presença de um domínio para a enzima transposase, a qual apresenta similaridades na sequência de aminoácidos com os elementos deste grupo (KEMPKEN; WINDHOFER, 2001). Em relação às TIRs, elas podem divergir bastante em suas sequências entre diferentes elementos (ARENSBURGER et al., 2011; RUBIN; LITHWICK; LEVY, 2001). Foram observados genes para a enzima transposase que apresentam quatro éxons preditos em peixes (KOGA et al., 2011) e cinco éxons em *Drosophila* (ROSSATO et al., 2014). Esses genes codificam uma sequência que apresenta em torno de 500 à 800 aminoácidos (LADEVÈZE et al., 2012; ROSSATO et al., 2014). As transposases de *hAT* apresentam uma estrutura proteica com quatro domínios: uma região N-terminal com domínio de dedos-de-zinco, um segundo domínio de ligação ao DNA, um domínio catalítico e um domínio de

inserção que contém numerosas  $\alpha$ -hélices (ARENSBURGER et al., 2011). Ainda há a presença de seis blocos conservados, sendo que sua característica mais conservada entre diferentes elementos é um domínio de aproximadamente 50 aminoácidos localizado na porção C-terminal (RUBIN; LITHWICK; LEVY, 2001).

A superfamília *hAT* agrupa elementos de três famílias, *Ac*, *Buster* (ARENSBURGER et al., 2011), e *Tip* (ROSSATO et al., 2014), onde a identificação é baseada na especificidade das TSDs geradas por elementos de cada família e na sequência das transposases, sendo que nesta última, as principais diferenças se encontram no domínio de ligação ao DNA e no domínio de inserção (ARENSBURGER et al., 2011; ROSSATO et al., 2014).

Os TEs do tipo *hAT* se movem pelo genoma por um mecanismo onde há quebra de dupla fita nas extremidades do TE através da formação de uma estrutura em grampo no DNA adjacente e, com a formação desta estrutura em grampo, a extremidade 3'OH do transposon é exposta, a qual se une ao DNA alvo para sua inserção (ZHOU et al., 2004). Ladevèze et al. (2012) observaram que altas taxas de transposição estão relacionadas com maior frequência no aparecimento de novas mutações decorrentes do processo de inserção, ou ainda excisão imprecisa destes elementos, sendo que estes fatores podem levar à ocorrência de alterações no genoma do hospedeiro. Glugoski et al. (2018) observaram que a invasão de elementos *hAT* no genoma de peixes do gênero *Rineloricaria* contribuíram para a dispersão do rDNA 5S e para a ocorrência de rearranjos cromossômicos. O envolvimento de TEs *hAT* com rDNA também foi observado em peixes da família Rachycentridae, havendo a colocalização de elementos *hAT* com o rDNA 18S (COSTA et al., 2013). Também foram encontrados elementos *hAT* envolvidos com rearranjos cromossômicos em *Drosophila*, onde estes elementos foram observados em regiões que sofreram deleções e inversões (LADEVÈZE et al., 2012). Já em fungos partes do gene da histona H3 foram capturados e mobilizados por *hAT*, gerando centenas de cópias nestes genomas, e levando a um grande aumento no tamanho dos genomas (DHILLON et al., 2019).

Rubin; Lithwick; Levy (2001) identificaram que uma grande parte das sequências de *hAT* encontradas em plantas, animais e fungos são sequências denominadas de “fósseis genéticos”, ou seja, que tiveram sua inserção nos genomas há muito tempo e já não apresentam atividade de transposição. Porém, nesse mesmo estudo também foram identificadas poucas sequências *hAT* com a presença das sequências TIRs conservadas, sugerindo que eles podem ainda estar ativos no genoma, ou terem apresentado período de reativação recente (RUBIN; LITHWICK; LEVY, 2001). Elementos *hAT* ativos já foram inclusive observados como causadores de alterações fenotípicas em peixes da espécie *Oryzias latipes*, ao se inserirem na

região codificante de um gene (HORI et al., 1998). O elemento *Tol2* também apresenta alta similaridade nas sequências nucleotídicas encontradas nas suas cópias dentro do genoma de espécies de *Oryzias*, o que levou a proposição de uma invasão recente deste elemento neste genoma (KOGA et al., 2000). *hAT* ativos também foram encontrados nos genomas de diferentes espécies de *Drosophila*, demonstrando inúmeras sequências para transposase intactas (ORTIZ; LORETO, 2009).

Os elementos *hAT*, junto com os elementos *Tc1-Mariner*, correspondem aos transposons mais abundantes no genoma de algumas espécies de peixes, demonstrando o importante papel que estes TEs têm dentro deste grupo (GAO et al., 2016; GARCÍA et al., 2019). Inclusive, em *Solea senegalensis*, os TEs *hAT* correspondem a cerca de 1% do genoma, sendo os TEs mais abundantes no genoma desta espécie (GARCÍA et al., 2019).

### 1.3 TEs na evolução do genoma

Os TEs possuem um “ciclo de vida”, descrito por Kidwell; Lisch (2001), o qual é composto por três fases: (i) uma fase de invasão, onde ocorre amplificação das sequências, aumentando seu número de cópias e o surgimento de algumas mutações, que fazem com que algumas sequências se deteriore; (ii) em seguida, uma fase de maturação, onde as taxas de aumento no número de cópias e perdas por mutações estão mais ou menos balanceadas e; (iii) a fase de senescência, que pode se estender por milhões de anos, onde não ocorre mais a amplificação, e essas sequências tendem a acumular mutações e se degenerarem. Esse processo de deterioração ocorre de diferentes maneiras de acordo com as diferentes estruturas e mecanismos de amplificação dos TEs e acaba levando a uma perda de identidade destes elementos (FERNÁNDEZ-MEDINA et al., 2012). Durante essa etapa de degeneração é possível que essas sequências sirvam como matéria de base para que as mutações aleatórias promovam inúmeras modificações, as quais poderiam levar essas sequências assumirem novas funções genômicas, o que caracteriza o mecanismo denominado de coopção molecular (MILLER; MCDONALD; PINSKER, 1997). Portanto, os TEs podem influenciar na evolução do genoma de três maneiras diferentes: através de alterações nas funções gênicas pela sua inserção, através da indução de rearranjos cromossômicos e, como matéria de base para a evolução de novos genes ou sequências regulatórias (FESCHOTTE; PRITHAM, 2007).

A forma mais simples em que os TEs podem alterar o genoma é através de suas inúmeras transposições, gerando aumento no tamanho e na diversidade do genoma (WARREN et al., 2015). Em adição, entre diferentes grupos de espécies há uma grande variação nas quantidades

de TEs presentes em cada genoma, bem como na diversidade de classes e famílias de elementos presentes (WARREN et al., 2015). Além do aumento no tamanho do genoma, as inúmeras transposições dos TEs podem causar alterações na expressão de genes, sendo que a inserção de um TE pode ocorrer em regiões gênicas dos genomas, podendo assim causar disfunções genéticas (CHÉNAIS et al., 2012). Ainda em relação a alterações causadas pela movimentação destes elementos, TEs do grupo dos *Helitrons* apresentam importante influência na evolução genômica, pois frequentemente capturam e mobilizam genes, os quais podem ser destruídos por acúmulo de mutações se não oferecerem vantagens para o hospedeiro, ou podem ser mantidos se apresentarem alguma vantagem (KAPITONOV; JURKA, 2001, 2007; THOMAS; PRITHAM, 2015). Portanto, a captura de genes pode contribuir para a geração de novos produtos gênicos para o indivíduo, fazendo com que este elemento tenha potencial de gerar variações no genoma ao longo da evolução (GRABUNDZIJA et al., 2016).

Outros estudos ainda demonstraram a atuação dos TEs na origem de rearranjos cromossômicos, gerando variações estruturais entre cromossomos (BOURQUE et al., 2018; EICHLER; SANKOFF, 2003; GLUGOSKI et al., 2018; LONGO et al., 2009). Variações estruturais como deleções, duplicações, inversões e translocações podem ser geradas através de recombinação homóloga ou ectópica, sendo que, na escala evolutiva, estes rearranjos mediados por TEs são um importante mecanismo que pode levar à especiação (WARREN et al., 2015).

Além destes casos, inúmeros casos e processos já foram descritos onde TEs degenerados assumem novas funções para o genoma do hospedeiro, podendo executar papéis de regulação da expressão, tanto transcricional quanto pós-transcricional (FESCHOTTE, 2008). Os TEs podem atuar fornecendo material de base para o surgimento de elementos regulatórios, gerar genes para microRNAs, ou ainda, mutações na sua sequência pode gerar informação para outras proteínas que podem ser funcionais para o organismo, entre outros (FESCHOTTE, 2008). Sequências de TEs não autônomos (MITes), geralmente presentes em genomas eucarióticos em grandes quantidades, também são responsáveis por importantes alterações nos genomas dos hospedeiros, os quais já foram observados atuando na regulação da expressão gênica e na geração de microRNAs e lncRNAs (MENZEL et al., 2014; PERUMAL et al., 2020; WANG et al., 2016).

As inserções de elementos transponíveis que são mantidas e se multiplicam ao longo da evolução geralmente ocorrem em regiões denominadas “refúgios seguros”, que são regiões no genoma onde alterações são mais toleradas pelo organismo (WERREN, 2011). Nestas regiões eles não apresentam efeitos negativos significativos para o genoma e tem menos chances de

serem eliminados ou silenciados e, conseqüentemente, esses elementos podem ser “cooptados” pelo genoma (WERREN, 2011).

Sequências derivadas de TEs são observadas em muitas regiões promotoras e possuem atuação na regulação da expressão destes genes (FESCHOTTE, 2008; WERREN, 2011). Pesquisadores observaram que cerca de 25% dos promotores gênicos do genoma humano contém sequências derivadas de TEs (JORDAN et al., 2003). Em uma pesquisa realizada por van de Lagemaat et al. (2003), a partir de análises nos genomas de humanos e camundongos, os pesquisadores determinaram que genes do sistema imune, resposta a estresse e resposta a estímulos externos apresentam altas quantidades de TEs em suas proximidades, os quais atuam na regulação da atividade destes genes. O oposto é observado em genes altamente conservados, como aqueles relacionados ao metabolismo ou estrutura celular, onde alterações em sua expressão seriam menos toleradas (van de LAGEMAAT et al., 2003). A coopção de proteínas derivadas de TEs é um mecanismo de grande importância para a evolução dos sistemas de defesa, onde proteínas derivadas de TEs assumiram importantes funções contra a invasão e proliferação de agentes infecciosos, como vírus e até contra outros elementos transponíveis (JANGAM; FESCHOTTE; BETRÁN, 2017). Inclusive, foi descoberto que a origem de proteínas do sistema de recombinação V(D)J, presente em células do sistema imune de vertebrados, ocorreu a partir de TEs da superfamília *Transib* (KAPITONOV; JURKA, 2005).

Entre outros estudos envolvendo a coopção molecular de TEs destacam-se a descoberta de que genes sincitina, importante proteína para o desenvolvimento da placenta e sobrevivência do embrião, se originaram a partir da coopção molecular do gene para o envelope proteico (*env*) presente em retrotransposons do tipo ERVs (*endogenous retroviruses*), sendo este evento de grande importância para a origem dos mamíferos placentários (LAVIALLE et al., 2013). Além destas novidades evolutivas, em *Drosophila* foi visualizado que elementos *Helitron* forneceram sequências para sítios de ligação necessários na formação do mecanismo de compensação de dose (ELLISON; BACHTROG, 2013).

Em peixes, Gao et al. (2020) encontraram casos em que TEs degenerados assumiram novas funções, sendo descobertos quatro genes funcionais derivados de TEs da superfamília *Tc1-Mariner*, além de outros pseudogenes presentes em múltiplas cópias. Em peixes da família Parodontidae, Traldi et al. (2019) observaram a presença de uma sequência com similaridade ao retrotransposon ERV compondo parte da ORF para histona H1, e tendo em vista a degeneração desta sequência, foi sugerida a possibilidade da coopção molecular deste elemento para geração de histonas variantes.

### 1.3.1 Atuação dos elementos transponíveis na evolução dos cromossomos sexuais

É proposto que o primeiro passo para diferenciação dos cromossomos sexuais, inicialmente homólogos (proto-cromossomos sexuais), seria a restrição da recombinação entre a região que abriga o(s) gene(s) determinante(s) do sexo, processo geralmente desencadeado por um acúmulo de sequências repetitivas nesta região de restrição à recombinação (CHARLESWORTH, 1991). Segundo Steinemann e Steinemann (1998), o acúmulo de TEs estaria envolvido na primeira etapa de diferenciação dos cromossomos sexuais em uma série de organismos. Logo o aumento no conteúdo de DNA por acúmulo de sequências repetitivas seria uma das primeiras mudanças a ocorrerem nos cromossomos sexuais em formação, sendo um sinal de que a recombinação é rara na região restrita ao cromossomo sexual heteromórfico (CHARLESWORTH; CHARLESWORTH; MARAIS, 2005). O principal efeito desta falta de recombinação seria então a instabilidade do cromossomo W/Y e sua degeneração, causando a perda da maior parte da atividade gênica (ŚLIWIŃSKA; MARTYKA; TRYJANOWSKI, 2016). A importância da atuação dos elementos repetitivos na evolução dos cromossomos sexuais é confirmada com a observação de que os cromossomos sexuais, tanto de plantas quanto de animais, apresentam diferentes tipos de sequências de DNAs repetitivos e em maiores quantidades em comparação com os autossomos (BISCOTTI; OLMO; HESLOP-HARRISON, 2015).

Segundo Śliwińska, Martyka e Tryjanowski (2016), a inserção e acúmulo dos TEs nos cromossomos sexuais ocorre na vizinhança de genes ligados ao sexo e, ao longo da evolução, essas sequências podem adquirir funções de mediar mecanismos de regulação da atividade gênica nestes cromossomos. Como já mencionado, além de atuar em funções regulatórias, os TEs ainda podem induzir rearranjos cromossômicos, como inversões, deleções e translocações (CHALOPIN et al., 2015a). Logo, tais rearranjos podem ser importantes para a evolução dos cromossomos sexuais ao gerar heteromorfismo entre os cromossomos X/Y ou Z/W (CHARLESWORTH; CHARLESWORTH; MARAIS, 2005).

Em um estudo a respeito da distribuição de TEs em cromossomos sexuais de peixes foi proposto que a abundância de TEs em regiões determinantes do sexo indica que estes elementos podem possuir sítios de ligação para fatores de transcrição em genes envolvidos no desenvolvimento gonadal ou, possuir envolvimento com mecanismos da compensação de dose para os cromossomos sexuais (CHALOPIN et al., 2015a). Logo, essas sequências podem influenciar o curso da cascata de expressão gênica para a determinação do sexo doando novos

elementos regulatórios (HERPIN et al., 2010; SCHEMBERGER et al., 2016). Em *Oryzias latipes* foi detectada uma inserção de partes de TEs na região *upstream* do gene *DmY* (gene duplicado a partir do *dmrt1* e presente no cromossomo Y), sendo que estes TEs alteraram a possibilidade de regulação deste que foi considerado o principal gene para determinação do sexo em machos desta espécie (HERPIN et al., 2010). Para esta mesma espécie foi detectado outro caso de inserção de um TE atuando na regulação do *dmrt1* (presente nos autossomos), e neste caso, a inserção do elemento *Rex1* contribuiu gerando sítios de ligação de um fator de transcrição Sox5, o qual atua regulando a expressão de *dmrt1* (SCHARTL et al., 2018).

Os peixes teleósteos apresentam uma ampla variedade de sistemas de cromossomos sexuais, os quais foram gerados através de múltiplos eventos independentes na evolução das linhagens, o que faz com que este seja um dos grupos mais interessantes para o estudo da evolução dos cromossomos sexuais e da determinação do sexo (CHALOPIN et al., 2015a). Parodontidae constitui um grupo de peixes Neotropicais que apresentam grande diversificação de sistemas de cromossomos sexuais, incluindo espécies com ausência de cromossomos sexuais diferenciados, com proto-cromossomos sexuais, com cromossomos sexuais diferenciados do tipo simples ZZ/ZW, ou múltiplo do tipo ZZ/ZW<sub>1</sub>W<sub>2</sub> (SCHEMBERGER et al., 2011).

Ziemniczak et al. (2014) demonstraram que o cromossomo W de Parodontidae sofreu diferenciação e degeneração de sua atividade gênica através do acúmulo de sequências repetitivas de diferentes tipos e classes. Schemberger et al. (2014) isolaram uma biblioteca C<sub>0</sub>t-1 para DNA repetitivo em Parodontidae e verificaram um intenso acúmulo de sequências repetitivas satélites e de elementos transponíveis, entre as quais, algumas foram relacionadas a degeneração molecular do cromossomo W. Os TEs *Tc1-Mariner* também foram localizados na região específica do cromossomo W (REW) e os dados de sequenciamento de DNA identificaram inúmeras cópias MITEs de elementos *Tc1-Mariner* em Parodontidae (SCHEMBERGER et al., 2016).

### 1.3.2 Estudos moleculares em *Apareiodon* sp.

Recentemente, sequenciamentos de nova geração foram obtidos para *Apareiodon* sp. (detentora de sistema de cromossomos sexuais ZZ/ZW). Estes dados resultaram na montagem dos genomas macho (42 x cobertura) e fêmea (47 x de cobertura) com objetivos de anotação da fração repetitiva do genoma (SCHEMBERGER et al., 2019). Utilizando a ferramenta RepeatMasker foi demonstrado que ~36% do genoma de *Apareiodon* sp. é composto por DNAs repetitivos, os quais estão organizados em seus diferentes tipos (SCHEMBERGER et al., 2019).

A idade relativa dos elementos transponíveis nos genomas macho e fêmea foram estimadas utilizando os valores distâncias de *Kimura-2 parameters* (SMIT; HUBLEY; GREEN, 2015) e demonstraram dois períodos (fases) mais intensos de invasão de elementos transponíveis no genoma de *Apareiodon* (SCHEMBERGER et al., 2019).

Utilizando a ferramenta RepeatExplorer foi demonstrada uma comparação do número de cópias de elementos repetitivos entre os genomas macho e fêmea. Entre os dados obtidos de 69 clusters, três clusters tiveram maior proporção no genoma fêmea (SCHEMBERGER et al., 2019). A análise destes clusters demonstrou que provavelmente os TEs *DNA/CMC Enspm*, *DNA/hAT*, *DNA/Helitron*, *DNA/Tc1-Mariner* e *LTR/Gypsy* estão enriquecidos no genoma fêmea e, estes foram então selecionados e submetidos à localização *in situ*. Entretanto, os dados combinados das análises com o RepeatMasker, RepeatExplorer e localização *in situ* demonstraram que a fase mais antiga de invasão genômica de TEs no genoma de *Apareiodon* apresentou as cópias de elementos *Tc1-Mariner*, *EnSpm* e *Helitron* e que foram responsáveis pelo aumento de tamanho e organização da REW (SCHEMBERGER et al., 2019). A taxa de substituição neutra por geração foi determinada para estes três transposons de DNA, as quais mostraram o mesmo período de invasão no genoma de *Apareiodon sp.*, estimado entre 20 e 12 milhões de anos atrás (SCHEMBERGER et al., 2019). Além destes, os dados genômicos demonstraram maior proporção de sequências microssatélites no genoma fêmea quando comparado ao genoma macho. Estes microssatélites também demonstraram acúmulos na REW de *Apareiodon sp.* (SCHEMBERGER et al., 2019). Com isso, estes autores propuseram que a fase de invasão antiga de TEs acumulou sequências restritas a um elemento do par cromossômico que culminou no evento de supressão de recombinação no par protosexual para origem do cromossomo sexual W em *Apareiodon sp.*, com estimativa de início a cerca de 20 milhões de anos atrás. Desde então, a diferenciação molecular e a erosão da atividade gênica do cromossomo W de *Apareiodon sp.* pode ter ocorrido por acúmulos de sequências repetitivas de diferentes tipos e classes, as quais podem formar domínios na cromatina e ter uma relação com o controle da expressão dos prováveis genes da determinação sexual.

A descrição do perfil de deterioração dos TEs é de grande importância para elucidar como é a distribuição e abundância destes elementos no genoma (FERNÁNDEZ-MEDINA et al., 2012). O conhecimento sobre o acúmulo de TEs em genomas pode fornecer informações sobre o desenvolvimento de características específicas dentro de um grupo de espécies e os mecanismos envolvidos na determinação destas características (PLATT; BLANCO-BERDUGO; RAY, 2016). Portanto, à medida que novos genomas se tornam disponíveis, é

possível uma ampliação dos conhecimentos a respeito da contribuição de diferentes TEs para características específicas de cada genoma e sua evolução (SOTERO-CAIO et al., 2017).

## 2. JUSTIFICATIVA E OBJETIVOS

Elementos transponíveis são sequências que estão envolvidas com frequência em modificações genômicas e cariotípicas em eucariotos, conseqüentemente apresentando grande influência na composição e evolução dos genomas destas espécies. Sabe-se que os genomas de peixes apresentam uma grande diversidade cariotípica, mesmo entre espécies da mesma família. Nesse grupo ocorre diferentes sistemas de cromossomos sexuais sujeitos as modificações impostas pelos TEs, além da maioria das espécies com dados genômicos apresentarem grande diversidade de tipos de TEs e maior abundância de transposons de DNA. Estes TEs, até o presente momento, não foram amplamente estudados para este grupo, principalmente em espécies detentoras de sistemas de cromossomos sexuais ZZ/ZW e em grupos de espécies de peixes Neotropicais. Assim, tendo em vista a importância evolutiva destes elementos e a escassez de estudos a respeito destas sequências voltados para peixes, o objetivo geral deste trabalho foi a caracterização dos TEs da superfamília *hAT* nos genomas macho e fêmea de *Apareiodon* sp., bem como, a investigação a respeito da autonomia de sequências, degeneração molecular, presença nos cromossomos sexuais, manutenção como sequências neutras no genoma ou, na proposição de mecanismos de origem de DNAs satélites e coopção de sequências.

Os objetivos específicos deste trabalho foram:

- Prospecção e caracterização molecular de sequências *hAT* nos genomas macho (ZZ) e fêmea (ZW) de *Apareiodon* sp.;
- A identificação da presença de sequências e domínios funcionais e/ou degenerados nas sequências das transposases de *hAT*;
- A identificação de sequências de TIRs e TSDs flanqueando as sequências das transposases;
- A construção de agrupamentos de sequências de aminoácidos conservadas de *hAT* de *Apareiodon* sp. com elementos de famílias conhecidas na literatura para sua correta identificação;
- Estimativa do tempo de inserção e da conseqüente degeneração molecular destes elementos transponíveis no genoma de *Apareiodon* sp.;
- Identificação e caracterização de sequências MITEs de *hAT*;

- Avaliação da distribuição de sequências repetidas *in tandem* nos contigs com a presença de elementos *hAT* e identificação do possível acúmulo de sequências repetidas *in tandem* no interior destes TEs;
- Avaliação da distribuição cromossômica de TEs pertencentes a superfamília *hAT* no cariótipo de *Apareiodon* sp. com atenção especial àqueles localizados nos cromossomos Z e W.

### 3. MATERIAL E MÉTODOS

#### 3.1 Seleção e caracterização das sequências

A identificação e análise das sequências dos TEs da superfamília *hAT* foi realizada a partir da informação da fração repetitiva do genoma de *Apareiodon* sp., onde duas bibliotecas (macho e fêmea) haviam sido previamente sequenciadas por Schemberger et al. (2019) por meio da plataforma Illumina HiSeq e anotadas com a utilização do software RepeatMasker (SMITH; HUBLEY; GREEN, 2013-2015) e de uma biblioteca customizada de sequências repetitivas obtida com o RepeatModeler (SMIT; HUBLEY, 2008–2015). No presente trabalho, para a seleção de sequências dos TEs *hAT* na fração repetitiva foi realizada a anotação destes elementos nos genomas macho e fêmea de *Apareiodon* sp. com a utilização do software RepeatMasker (SMITH; HUBLEY; GREEN, 2013-2015). A anotação destas sequências em ambos os genomas foi realizada utilizando uma biblioteca de TEs customizada, a qual foi obtida a partir de TEs autônomos ou mais conservados de peixes presentes nos bancos de dados Repbase (KOHANY et al., 2006) e FishTEDB (SHAO et al., 2018). Essas sequências conservadas de *hAT* em peixes foram utilizadas para a obtenção de domínios proteicos característicos deste TE utilizando a ferramenta PFAM (EL-GEBALI et al., 2019). Nesta etapa foi possível prospectar três domínios proteicos característicos de *hAT* presentes em peixes (zf-BED, Dimer Tnp hAT e DBD Tnp Hermes). Para a construção desta biblioteca foi aplicada uma metodologia similar à de Gao et al. (2016), que consistiu no uso das sequências de aminoácidos como query para tBLASTn contra o genoma de *Apareiodon* sp. Sequências satélite foram mascaradas na biblioteca e foi realizada uma filtragem para obtenção de sequências com tamanho maior que 80 pb e mais de 60% de similaridade à TEs da superfamília *hAT* para ambos os genomas macho e fêmea.

Em seguida, com a utilização das sequências obtidas após as etapas de filtragem para remoção de sequências não características de *hAT* e daquelas que detinham apenas repetições *in tandem* foi realizada a identificação e caracterização dos domínios proteicos pertencentes à *hAT* de *Apareiodon* sp. As sequências selecionadas foram submetidas à identificação e extração de suas ORFs com a utilização do software getorf (EMBOSS version 6.6.0.0). Posteriormente as sequências das ORFs obtidas foram utilizadas para a busca de domínios proteicos pertencentes à *hAT*. Para isso, essas sequências foram analisadas com o software HMMER (versão 3.1b1), utilizando os parâmetros padrão baseados no banco de dados atualizado do PFAM (EL-GEBALI et al., 2019). A partir da lista obtida, com todos os domínios detectados

nas sequências analisadas foram filtrados os resultados para a obtenção das sequências com a presença de um ou mais domínios pertencentes à transposase de *hAT*. Posteriormente ainda foi realizada a filtragem das sequências de acordo com as principais famílias observadas de acordo com a classificação do RepeatMasker, onde foi detectada uma grande abundância de quatro grupos diferentes: hAT-Ac, hAT-Charlie, hAT-Blackjack e hAT-Tip100.

Para aqueles TEs que apresentaram suas sequências de nucleotídeos maiores do que 1000 pb, assim tendo a possibilidade de apresentar uma transposase mais conservada, foram realizadas análises utilizando o software FGENESH (SOLOVYEV et al., 2006) buscando a determinação da ORF para o gene da transposase ou do número de éxons deste gene e a localização de sequências de início e término de transcrição, além do alinhamento das sequências de aminoácidos destas transposases com o software MUSCLE (EDGAR, 2004) para observação da conservação de resíduos de aminoácidos entre diferentes elementos.

Posteriormente, as sequências dos TEs foram analisadas para detecção de TIRs e TSDs. As TIRs foram identificadas pela busca manual de sequências entre 5 e 27 pb, com as características (T/C)A(A/G)NG, descritas por Rubin; Lithwick; Levy (2001) como a sequência consenso de TIRs de *hAT*. Para as TSDs foram então buscadas sequências de 8 pb adjacentes às TIRs com as características determinadas para cada família como descritas por Arensburger et al. (2011). Esta busca foi realizada entre as sequências que flanqueiam as porções 5' e 3' das sequências nas quais foram obtidos domínios da enzima transposase conservados.

### 3.2 Construção do Repeat Landscape

Foram construídos gráficos Repeat Landscape para as sequências de *hAT* analisadas no genoma macho e no genoma fêmea de *Apareiodon* sp. Para construção destes gráficos foram utilizadas as sequências previamente separadas nos quatro principais grupos de TEs *hAT* obtidos em *Apareiodon* sp. segundo a classificação do RepeatMasker (hAT-Ac, hAT-Charlie, hAT-Blackjack e hAT-Tip100).

Para a construção do gráfico contendo a taxa de substituição de Kimura das famílias *hAT* no genoma de *Apareiodon* sp. em relação ao tempo (*landscape*), as sequências de cada família (Ac, Blackjack, Charlie e Tip100) foram novamente submetidas ao programa RepeatMasker v. 4.0.1 (<http://www.repeatmasker.org/>) seguindo os parâmetros do desenvolvedor. Os resultados obtidos foram acurados pelos *scripts* calcDivergenceFromAlign.pl e createRepeatLandscape.pl que geraram as taxas de substituição de Kimura para cada família *hAT*. O tempo de inserção (T) das sequências foi calculado de

acordo com o descrito por Schemberger et al. (2019) aplicando uma taxa de substituição média ( $r$ ) de  $10^{-8}$  sítios de substituições sinônimas por ano característica de TEs (ARKHIPOVA; MESELSON, 2005) na fórmula  $T = k/2r$  (LI et al., 2019). Com isso, as taxas de substituição de Kimura em relação ao tempo de dispersão foram plotadas em um gráfico de linhas construído pelo pacote R ggplot2.

### 3.3 Agrupamentos de similaridade de sequências

As sequências similares à TEs *hAT* obtidas no genoma de *Apareiodon* sp. foram utilizadas para a análise da diversidade das diferentes famílias de *hAT* distribuídas no genoma desta espécie. Para esta análise foi realizada a seleção das sequências mais íntegras observadas no genoma fêmea de *Apareiodon* sp., as quais apresentaram todos os domínios da transposase, além de também terem sido selecionadas as sequências que apresentaram apenas o domínio C-terminal de dimerização (Dimer tnp *hAT*), para as quais foram filtradas apenas as sequências que apresentassem similaridade maior que 90% ao respectivo domínio de *hAT* e e-value menor que 0,01. Após, foram selecionadas sequências de TEs *hAT* identificados entre as principais famílias *hAT* presentes nos bancos de dados atualizados do Dfam (STORER et al., 2021), e do NCBI (<https://www.ncbi.nlm.nih.gov>) como identificadas por Arensburger et al. (2011) para serem utilizadas como sequências de referência no agrupamento com as sequências de *Apareiodon* sp. As respectivas sequências de referência utilizadas encontram-se listadas na tabela 1.

**Tabela 1:** Sequências de referência obtidas através dos bancos de dados do Dfam e do NCBI utilizadas para os agrupamentos de sequências *hAT* de *Apareiodon* sp. Indicados os números de acesso e IDs das sequências utilizadas, suas respectivas espécies e os grupos aos quais cada sequência pertence.

Nº acesso	ID	Espécie/Clado	Classificação
DF0002841	hAT-26 DR	<i>Danio rerio</i>	Ac
DF0002255	hAT-31 DR	<i>Danio rerio</i>	Ac
DF0003736	hAT-16 DR	<i>Danio rerio</i>	Ac
DF0000074	BLACKJACK	Eutheria	Blackjack
DF0003203	Chaplin1 DR	<i>Danio rerio</i>	Charlie
DF0002322	hAT-5 DR	<i>Danio rerio</i>	Charlie
DF0003698	hAT-12 DR	<i>Danio rerio</i>	Charlie
DF0002972	hAT-13 DR	<i>Danio rerio</i>	Tip100
DF0003856	hAT-23 DR	<i>Danio rerio</i>	Tip100
DF0001293	hAT-1_Mam	Monotremata	Tag1
AAC37217	Hermes	<i>Musca domestica</i>	Hermes

AFC96945	Tgf2	<i>Carassius auratus</i>	Tgf2
BAA87039	Tol2	<i>Oryzias latipes</i>	Tol2
DF0001645	HOBO	<i>Drosophila</i>	hobo
CAB68118	Daysleeper	<i>Arabidopsis thaliana</i>	Daysleeper
CAA38906	Tam3	<i>Antirrhinum majus</i>	Tam3
CAA93759	Restless	<i>Tolypocladium inflatum</i>	Restless
AAS21248	Herves	<i>Anopheles gambiae</i>	Herves
ABF20549	DrBuster1	<i>Danio rerio</i>	Buster
ABF20543	AeBuster1	<i>Aedes aegypti</i>	Buster
ABF20548	CsBuster	<i>Ciona savignyi</i>	Buster

Duas análises de agrupamentos de sequências foram realizadas para determinação da diversidade das sequências *hAT*, uma contendo as sequências ditas conservadas da transposase *hAT* em *Apareiodon* sp., e outro com as sequências do Dimer tnp *hAT*. Para isso, as sequências selecionadas foram submetidas ao software MEGA-X (KUMAR et al., 2018), onde as sequências de referência obtidas para cada família foram alinhadas junto com as sequências selecionadas de *Apareiodon* sp. através da ferramenta MUSCLE e os alinhamentos foram submetidos para a construção de árvores pelo método Maximum Likelihood com 1000 replicações.

#### 3.4 Análise de sequências MITES

As sequências *hAT* não autônomas MITES foram analisadas e contabilizadas no genoma de *Apareiodon* sp. através da utilização do software MITE Tracker (CRESCENTE et al., 2018), o qual realizou a prospecção de sequências MITES a partir dos genomas macho e fêmea de *Apareiodon* sp., buscando por sequências com estruturas características das sequências terminais destes elementos.

Os dados obtidos foram então filtrados para a obtenção de sequências de MITES correspondentes à transposons *hAT*. Para essa filtragem as sequências MITES obtidas foram submetidas as ferramentas RepeatMasking (SMITH; HUBLEY; GREEN, 2013-2015) e CENSOR (KOHANY et al., 2006), sendo então selecionadas as sequências que apresentassem similaridade à *hAT*, as quais foram então agrupadas e contabilizadas entre as principais famílias observadas. Foram então analisadas as sequências de TIRs presentes nestes elementos para determinação das sequências mais comuns para cada família.

#### 3.5 Avaliação da presença de repetições in tandem associadas à *hAT*

Foi realizado um levantamento de sequências repetidas *in tandem* contidas no interior das sequências *hAT*. As repetições *in tandem* foram anotadas nos genomas macho e fêmea e nos TEs a partir da utilização de uma biblioteca customizada de DNA satélite obtida com o software Repbase (KOHANY et al., 2006), junto com a utilização do software TRF (*Tandem Repeats Finder*) (BENSON, 1999). Assim foram selecionadas as sequências de repetição *in tandem* dos tipos mais representativos presentes no interior de TEs *hAT*, bem como, foram contabilizadas as frequências de cada tipo de repetição *in tandem* de forma comparativa nos genomas macho e fêmea de *Apareiodon* sp.

Também foram analisados os contigs do genoma de *Apareiodon* sp. com a presença de sequências *hAT* a respeito da abundância de repetições *in tandem* presentes associadas a estes TEs. Esta análise foi realizada através do software Dfam (STORER et al., 2021), onde os contigs foram submetidos para a prospecção de sequências de TEs e de repetições *in tandem* presentes nestas sequências. Em seguida foram selecionadas sequências com a presença de repetições *in tandem* flanqueadas por sequências similares a *hAT*.

### 3.6 Obtenção de cromossomos mitóticos

A obtenção de cromossomos mitóticos foi realizada de acordo com o método descrito por Bertollo, Cioffi e Moreira-Filho (2015), a partir do material renal da espécie *Apareiodon* sp., da família Parodontidae, detentora de sistema de cromossomos sexuais do tipo ZZ/ZW.

### 3.7 Obtenção das sequências e geração de sondas

Para o desenho dos primers, contigs do genoma de *Apareiodon* sp. contendo as sequências de *hAT* foram alinhados no software MUSCLE (EDGAR, 2004) para a detecção de regiões de degeneração mínima. A partir do alinhamento, foram desenhados primers para amplificação por PCR de sequências parciais do TE *hAT* com o uso do software Primer3plus (UNTERGASSER; NIJVEEN, 2007; ROZEN; SKALETSKY, 2000). Por meio destas análises foram obtidos dois conjuntos de primers para realização de PCR e obtenção de amplicons conforme descrito na tabela 2.

**Tabela 2:** Descrição dos primers, regiões alvo e tamanho de amplicons na amplificação de *hAT* no genoma de *Apareiodon* sp.

Alvo	Nome	Sequência	Amplicon
Domínio BED	HatFw_bed	5' GCAGCATTTTGGCTTCCTT 3'	~ 610 pb
	HatRv_bed	5' AGCGATGTTGACCCAGT 3'	
Domínio DIMER	HatFw_dimer	5' GAAGAGCGCCATACTCAACA 3'	~ 680 pb

	HatRv_dimer	5' TCAGGCAGAAGCAGAGACC 3'	
BED até DIMER	HatFw_bed HatRv_dimer	5' GCAGCATTTCCTTCCTT 3' 5' TCAGGCAGAAGCAGAGACC 3'	~ 1.900 pb

Posteriormente, o DNA genômico do fígado de exemplares de *Apareiodon* sp. foi extraído de acordo com o método brometo de cetil trimetilamônio (CTAB) (DOYLE; DOYLE, 1990). A amplificação de *hAT* por PCR utilizando os primers obtidos utilizou reação padrão contendo 40 ng de DNA genômico, 0,2 µM de cada primer, 0,4 mM de desoxinucleotídeos trifosfatos (dNTPs), 1× Taq tampão de reação (200 mM Tris, pH 8.4, 500 mM KCl), 2 mM MgCl<sub>2</sub> e 1 U de Taq DNA polimerase (Invitrogen). Os parâmetros da amplificação padrão foram: desnaturação inicial 10 min a 94 °C, 35 ciclos de três passos (1 min a 94 °C, 45 segundos a 60 °C e 2 min a 72 °C) e extensão final de 7 min a 72 °C. Os produtos da PCR foram verificados em gel de agarose 1%. Os amplicons resultantes foram purificados utilizando o *kit PCR DNA and Gel Band Purification* (GE Healthcare) e submetidos ao sequenciamento automático de DNA utilizando sequenciador de DNA *ABI-PRISM Genetic Analyzer* (Applied Biosystems). As sequências resultantes foram submetidas a confirmação de identidade dos segmentos obtidos nos softwares online CENSOR e DFam. Confirmada a identidade, as sequências foram marcadas com digoxigenina-11-dUTP utilizando o Digoxigenin Nick Translation Labeling Kit (Jena Bioscience) para geração de sondas e posterior uso em procedimento de localização *in situ*.

### 3.8 Localização *in situ*

A localização das sondas de *hAT* obtidas por PCR foi realizada pela técnica de hibridação *in situ* fluorescente (PINKEL; STRAUME; GRAY, 1986). A mistura de hibridação foi sintetizada com as seguintes condições: 2.5 ng/µL de sonda, 20% formamida, 5x SSC, e 10% sulfato dextrano. A hibridação ocorreu por 18 h à 42 °C. Banhos de pós-hibridização foram realizados nas condições: Triton X-100 0,3%/1xPBS em temperatura ambiente por 10 min, Triton X-100 0,1%/1xPBS em temperatura ambiente por 5 min. Para a detecção de fluorescência foi utilizado 10 µg/mL de anti-digoxigenin rhodamine (Roche Applied Science), a qual foi incubada a 37 °C por 1 hora no tampão de detecção 5% NFD/4% SSC. Os cromossomos foram contracolorados com 0,2 µg/mL de 4', 6-diamidino-2-fenilindole (DAPI) em meio de montagem Vectashield (Vector, Burlingame, CA) e analisados em microscópio de campo claro e epifluorescência Zeiss Axio Lab 1 acoplado ao software Zen Lite em Câmera Zeiss AxioCam ICM1, resolução de 1.4 megapixels (Carl Zeiss®).

### 3.9 Organização dos cariótipos

As melhores metáfases de fêmeas da espécie *Apareiodon* sp. foram fotografadas e os cromossomos organizados em cariótipos. Os cromossomos foram organizados em ordem decrescente de tamanho e de acordo com a morfologia seguindo a classificação de Levan et al. (1964), obedecendo a regra de razão de braços e organizados conforme o cariótipo de *Apareiodon* sp. descrito por Vicari et al. (2006).

## 4. RESULTADOS

### 4.1 Caracterização molecular do elemento *hAT*

Após as etapas de seleção e filtragem das sequências dos TEs *hAT* foram identificadas 88.721 e 84.171 sequências com similaridade à *hAT* nos genomas fêmea e macho de *Apareiodon* sp., respectivamente. Essas sequências representam um total de 3.661.770 pb dentro de genoma fêmea, e 3.646.091 pb no genoma macho, correspondendo a ~ 0,4% do genoma da espécie. Na biblioteca de fêmea as sequências *hAT* apresentam tamanhos que variaram de 80 – 3.976 pb, enquanto em macho estas variaram entre 80 – 3.963 pb.

As análises na biblioteca de fêmea demonstraram um total de 2.339 sequências que apresentaram um ou mais domínios proteicos pertencentes à transposase de *hAT*. Entre esses, foi possível detectar a ocorrência de três diferentes domínios, os quais são as três sequências pertencentes a transposons *hAT* identificadas pelo banco de dados do Pfam. Sendo estas, o domínio N-terminal de dedos-de-zinco (zf-BED), um segundo domínio de ligação ao DNA (DBD Tnp Hermes), e o domínio C-terminal de dimerização (Dimer Tnp *hAT*). Ainda, dentre estas sequências, apenas 11 apresentaram ambos os domínios C-terminal de dimerização e o domínio N-terminal de dedos-de-zinco, e apenas duas sequências apresentaram o domínio C-terminal de dimerização junto com o domínio de ligação ao DNA do elemento *Hermes*.

Entre estas sequências da biblioteca *hAT* do genoma fêmea com a presença de domínios proteicos pertencentes à transposase de *hAT* ainda foram selecionadas as sequências que apresentasse seus domínios preditos com valor de e-value <0,01, sendo encontradas 258 sequências para o domínio de dimerização C-terminal, o qual compreende a região mais conservada da transposase, 426 sequências para o domínio N-terminal de dedos de zinco, além de não terem sido encontradas sequências com a presença do domínio de ligação ao DNA relacionada à transposase de elementos da família *Hermes* com valor de e-value <0,01 (Tabela 3). Em relação a integridade dos domínios obtidos para as sequências, foram observados 358 domínios de dedos-de-zinco com integridade superior à 90% e para a região de dimerização foram 193 domínios com integridade superior à 90% (Tabela 3).

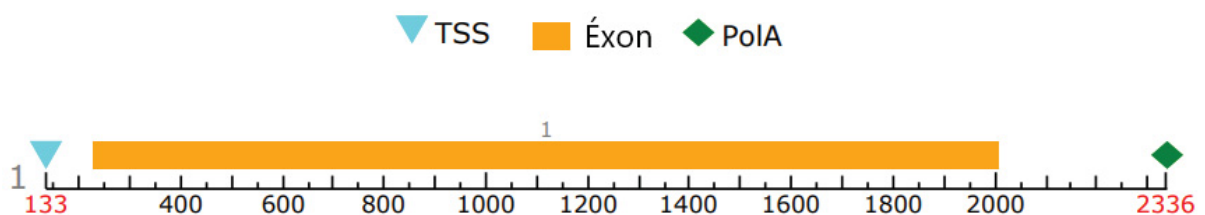
**Tabela 3:** Abundância dos domínios característicos de *hAT* presentes no banco de dados do Pfam nas sequências analisadas. Contagem dos três domínios característicos de *hAT* com integridade de suas sequências maior e menor do que 90% e e-value <0,01 no genoma fêmea.

Domínio	Integridade >90%	Integridade <90%
zf-BED	358	68
Dimer tnp <i>hAT</i>	193	65
DBD tnp Hermes	0	0

As sequências totais dos genomas macho e fêmea ainda foram separadas de acordo com sua semelhança a determinadas famílias descritas para *hAT*. A lista das sequências *hAT* obtida após as etapas de seleção e filtragem das sequências demonstrou uma grande abundância de sequências semelhantes a elementos pertencentes a quatro famílias dos elementos *hAT*, sendo estas: *hAT*-Ac, *hAT*-Blackjack, *hAT*-Charlie e *hAT*-Tip100. Foram obtidas 51.878 sequências no genoma fêmea e 48.476 no genoma macho pertencentes ao grupo *hAT*-Ac, o qual foi o mais numeroso, seguido por 14.832 sequências em fêmea e 14.562 sequências em macho pertencentes ao grupo *hAT*-Charlie. Para os elementos *hAT*-Blackjack foram obtidas 7.358 sequências em fêmea e 7.254 em macho, e para *hAT*-Tip100, a família menos numerosa, foram observadas 2.250 sequências em fêmea e 2.211 em macho.

#### 4.2 Análise da presença de elementos *hAT* íntegros em *Apareiodon* sp.

Para a avaliação da presença de elementos *hAT* íntegros em *Apareiodon* sp. foram selecionadas nove sequências, as quais apresentaram tamanho maior que 1.000 pb juntamente com domínios proteicos característicos de transposases *hAT*. Estas sequências foram analisadas com uso do software FGENESH para a detecção de possíveis ORFs para estes elementos e da presença ou ausência de íntrons, onde foram observadas ORFs para a enzima transposase de *hAT* com a presença de um único éxon (Fig. 1).



**Figura 1:** Representação da região codificante da sequência de nucleotídeos da transposase de TEs *hAT* putativamente íntegros no genoma de *Apareiodon* sp. Indicados na legenda o sítio de início da transcrição (TSS), éxon, e sinal de poliadenilação (PoliA).

As respectivas sequências de aminoácidos destes nove elementos analisados foram então alinhadas para análises de suas sequências da transposase, onde foi determinado se havia ou não a presença de transposases que apresentassem todos os seus domínios proteicos íntegros. A presença de ORF para a enzima transposase de *hAT* apresentando seus quatro domínios proteicos característicos descritos na literatura (Fig. 2A) pode ser observada para as nove sequências analisadas (Fig. 2B). Essas sequências que detinham os quatro domínios da transposase apresentaram um tamanho aproximado de 600 resíduos de aminoácidos. O alinhamento destas sequências demonstrou um alto nível de variabilidade (Fig. 2B), onde foram poucos os aminoácidos comuns entre elas, inclusive uma das sequências (hAT-3\_AP) apresentou diversas inserções não características de *hAT* na sua sequência.



abaixo das sequências. Os resíduos de aminoácidos conservados entre todas as sequências estão indicados por (\*), e os resíduos altamente conservados entre as transposases de *hAT*, responsáveis por funções catalíticas ou de ligação às TIRs apresentam-se destacados em verde.

No entanto, dentre estas sequências, cinco apresentaram todos os resíduos de aminoácidos necessários para a transposição (sendo denominadas hAT-1\_AP, hAT-4\_AP, hAT-5\_AP, hAT-6\_AP e hAT-7\_AP), os quais são altamente conservados entre transposases de diferentes TEs *hAT*, sendo responsáveis por reações catalíticas e de ligação às TIRs (Fig. 3).



**Figura 3:** Alinhamento das cinco sequências mais conservadas das transposases *hAT* de *Apareiodon* sp., as quais apresentaram todos os resíduos de aminoácidos necessários para a atividade de transposição de acordo com Atkinson (2014) e Yuan e Wessler (2011) conservados. Os quatro domínios característicos da transposase estão indicados abaixo das sequências. Os resíduos de aminoácidos conservados entre todas as sequências estão indicados por (\*), e os resíduos altamente conservados entre as transposases de *hAT*, responsáveis por funções catalíticas ou de ligação às TIRs apresentam-se destacados em verde.

As nove sequências previamente utilizadas, que apresentaram a sua sequência da transposase mais conservada, ainda foram utilizadas na busca de sequências em suas extremidades 5' e 3' com características de TIRs e TSDs de elementos *hAT*. Foram encontradas TIRs 5' e 3' similares àquelas descritas para elementos *hAT* em todas as cinco sequências que

apresentaram maior conservação, as quais detém tamanhos de 8 à 18 pb (Tabela 4). Já sequências TSDs foram encontradas adjacentes às TIRs em apenas três destas sequências, e apenas duas apresentaram TSDs íntegras, com todas as bases conservadas tanto na porção 5' quanto na 3' (Tabela 4).

**Tabela 4:** Sequências TIRs e TSDs encontradas nos TEs *hAT* de *Apareiodon* sp. Sequências das TIRs 5' e 3' flanqueando as sequências das transposases putativamente íntegras de cinco sequências *hAT* e sequências TSD observadas adjacentes às TIRs. Destacadas em negrito as bases conservadas tanto na extremidade 5' quanto na extremidade 3'.

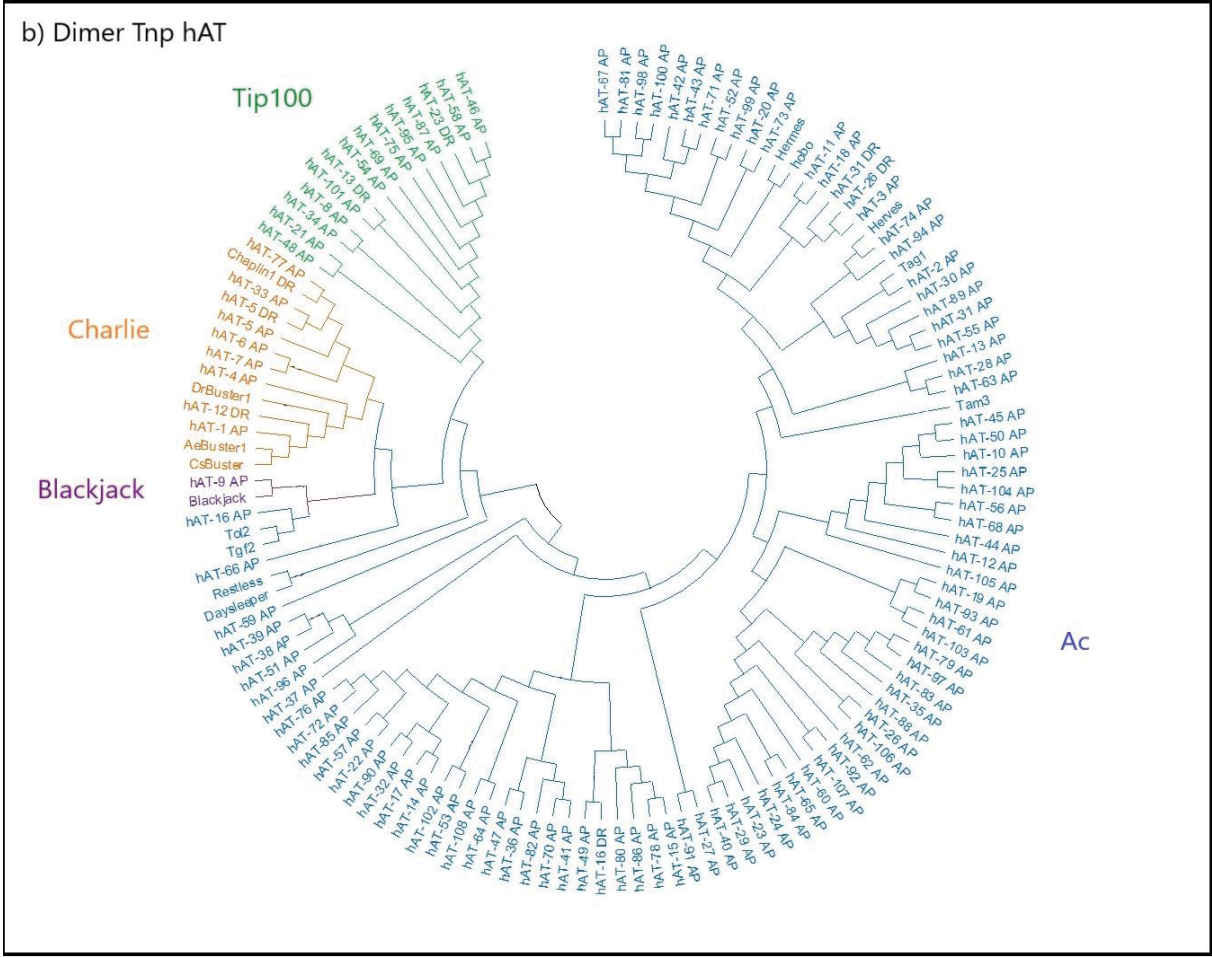
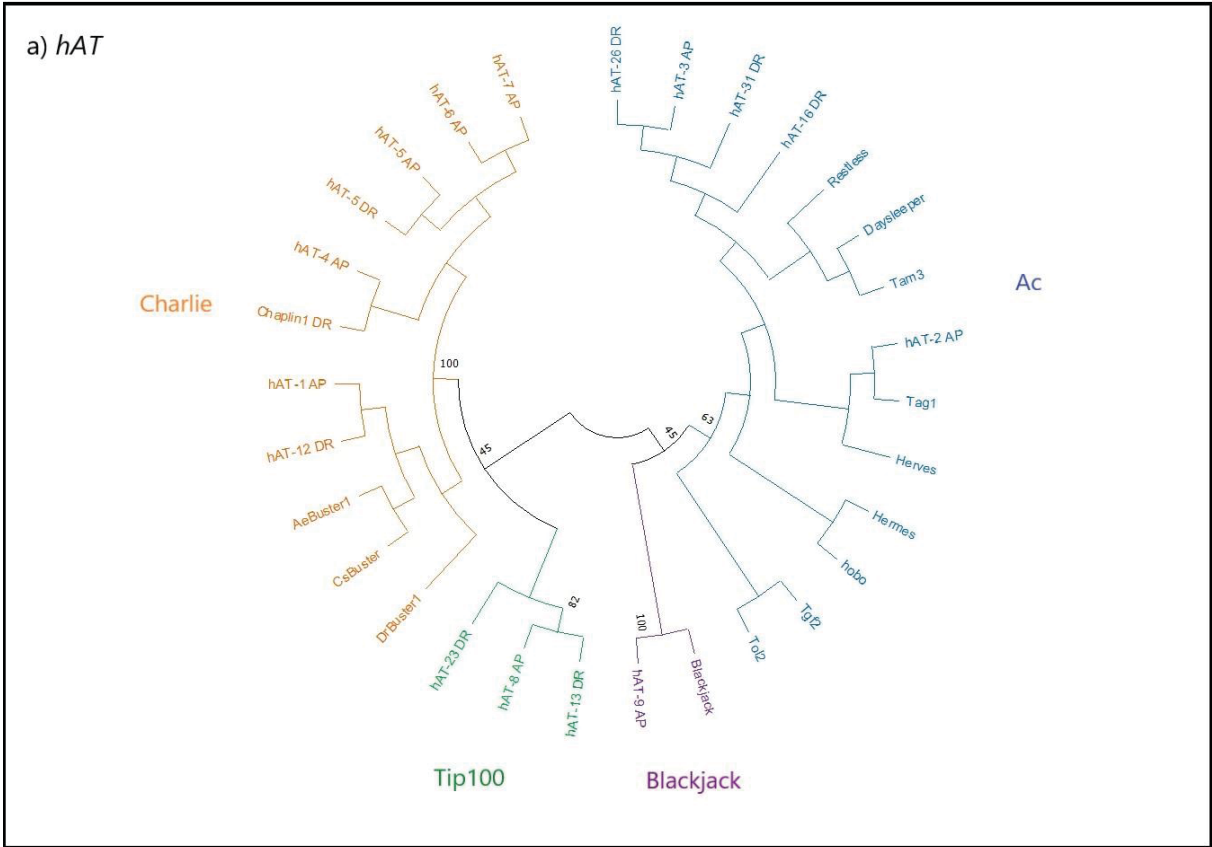
Sequência	TSD 5'	TIR 5'	TIR 3'	TSD 3'
<b>hAT-1_AP</b>	<b>CTCTAGAC</b>	<b>CAGCGATTCTCAACTGGT</b> ----- <i>Transposase</i> -----	<b>ACCAGTTGAGAAACCCCTG</b>	<b>CTCTAGAC</b>
<b>hAT-6_AP</b>	<b>TTCTAGAG</b>	<b>CAGGGGTCTCTA</b> ----- <i>Transposase</i> -----	<b>TTGAGACCCCTG</b>	<b>TTCTAGAG</b>
<b>hAT-7_AP</b>	<b>ATGTAGGA</b>	<b>CAAGGGTGGGCAAACT</b> ----- <i>Transposase</i> -----	<b>AGTTGCCCCACCCCTG</b>	<b>CTTTAAAC</b>
<b>hAT-4_AP</b>	-	<b>CAGTGTGA</b> ----- <i>Transposase</i> -----	<b>TCACACTG</b>	-
<b>hAT-5_AP</b>	-	<b>TAGTGTGA</b> ----- <i>Transposase</i> -----	<b>TCACACTA</b>	-

Ao fim destas análises foram obtidas cinco sequências que apresentaram todas as características necessárias para serem elementos íntegros e possivelmente autônomos: hAT-1\_AP, hAT-4\_AP, hAT-5\_AP, hAT-6\_AP e hAT-7AP. Estas sequências, putativamente íntegras, apresentaram tamanho aproximado de pouco mais de 2.000 pb, com sequências de TIRs em suas extremidades e uma ORF interna com todos os domínios da enzima transposase contendo cerca de 600 resíduos de aminoácidos.

#### *4.3 Agrupamentos de similaridade e identificação de diversidade de sequências*

A construção das árvores ML com as sequências de aminoácidos das transposases dos elementos *hAT* revelou a distribuição destes elementos nas quatro famílias mais abundantes previamente descritas para *Apareiodon* sp. (Fig. 4). Para o agrupamento com os elementos que apresentaram suas sequências transposase íntegras foi obtida uma maior abundância de TEs *hAT* pertencentes à família Charlie, sendo estas cinco sequências entre as nove utilizadas para o agrupamento (Fig. 4A). Para os outros grupos foram obtidas duas sequências pertencentes ao grupo Ac, sendo uma sequência (hAT-2\_AP) com similaridade à família Tag1, enquanto os grupos Blackjack e Tip100 exibiram apenas uma sequência cada (Fig. 4A).

Já em relação ao agrupamento com as sequências que apresentaram domínios C-terminais de dimerização (Dimer Tnp hAT) conservados foi observada uma grande abundância de sequências pertencentes ao grupo Ac, distribuídas em uma maior diversidade de famílias (Fig. 4B). Para os outros grupos foi obtido um número menor de sequências pertencentes aos grupos Charlie e Tip100, sendo Blackjack o grupo menos abundante (Fig. 4B).

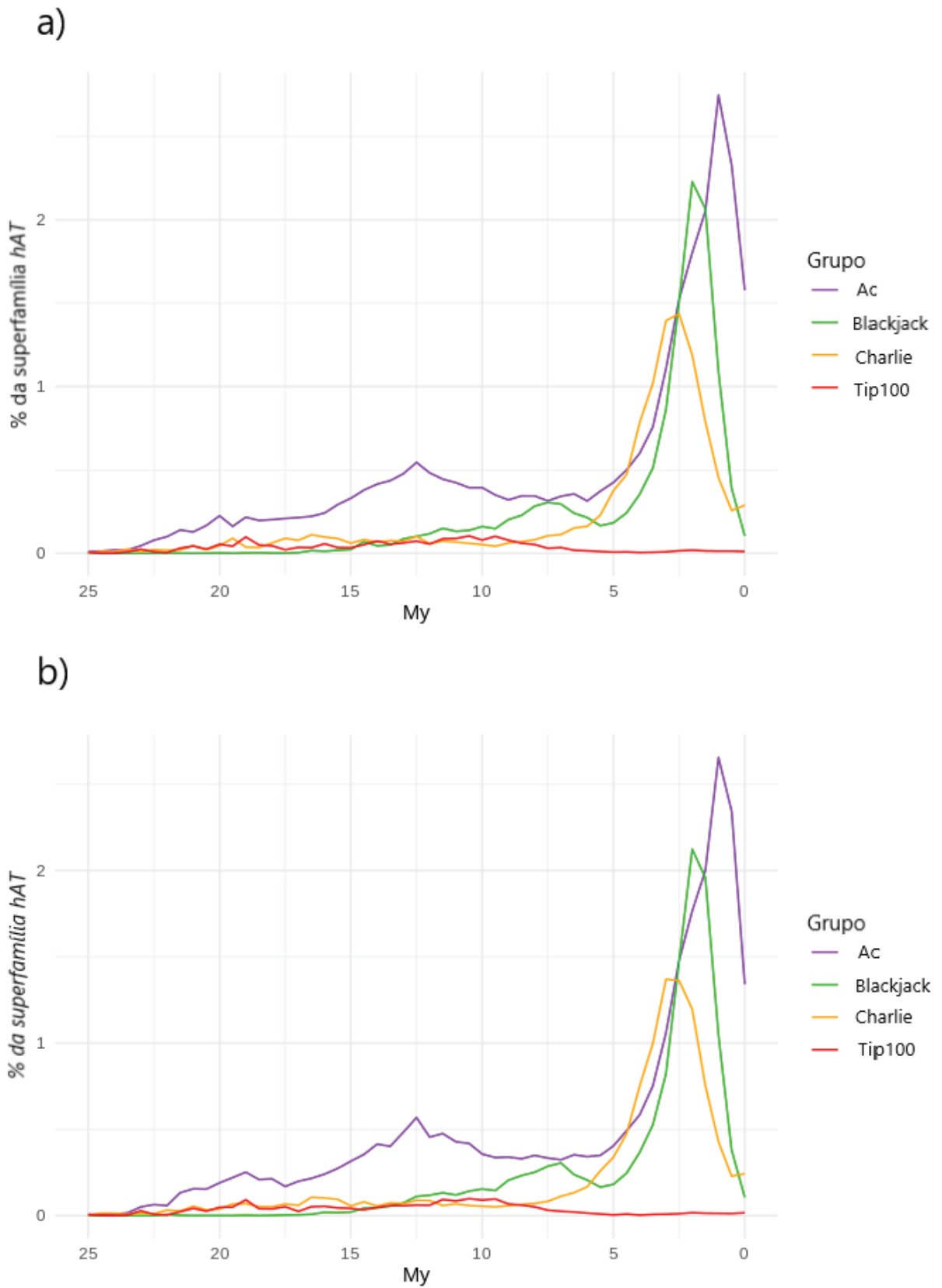


**Figura 4** – Agrupamentos de sequências utilizando o método Maximum Likelihood (ML) com as sequências de aminoácidos das transposases *hAT* de *Apareiodon* sp. e sequências de referência pertencentes às principais famílias identificadas para elementos *hAT*. Em a) agrupamento das nove sequências ditas conservadas e que apresentaram todos os domínios pertencentes a *hAT* em *Apareiodon* sp. Em b) agrupamento das sequências com a presença de domínios de dimerização C-terminais conservados de *Apareiodon* sp. Os quatro grupos principais obtidos (Ac, Blackjack, Charlie e Tip100) foram representados em cores nos ramos dos agrupamentos.

#### 4.4 Análise da invasão do TE *hAT* no genoma

Para a análise a respeito da invasão dos elementos *hAT* no genoma de *Apareiodon* sp. as sequências *hAT* totais dos genomas macho e fêmea, separadas nas principais subfamílias encontradas (Ac, Charlie, Blackjack e Tip100), foram utilizadas para a construção dos gráficos do Repeat Landscape.

Análises dos gráficos de substituição nucleotídica não demonstraram grandes diferenças de distribuição das sequências *hAT* nos genomas macho e fêmea de *Apareiodon* sp. (Fig. 5). A análise dos gráficos ainda demonstrou que as cópias *hAT* começaram a se expandir no genoma de *Apareiodon* sp. em um período de aproximadamente 20 milhões de anos atrás, com duas principais ondas de invasão no genoma, sendo uma primeira menor em um período entre 20 e 12 milhões de anos atrás, com uma maior abundância de sequências do grupo Ac, e uma segunda onda de invasão maior mais recente (a menos de 5 milhões de anos atrás) com maior abundância de sequências Ac, Blackjack e Charlie respectivamente, e com a ausência de novas inserções Tip100 (Fig. 5).



**Figura 5:** Estimativa de abundância e do tempo de inserção dos elementos correspondentes às famílias Ac, Blackjack, Charlie e Tip100 (superfamília *hAT*) em *Apareiodon* sp. Os picos indicam as ondas de invasão dos diferentes grupos de *hAT* nos genomas macho (a) e fêmea (b)

de *Apareiodon* sp. O eixo X indica estimativa do período de invasão dos elementos nos genomas (em milhões de anos), enquanto o eixo Y indica a abundância (%) destas sequências no conjunto total de sequências *hAT* presentes no genoma de *Apareiodon* sp.

#### 4.5 Análise de sequências MITEs

Com a análise das sequências MITEs pertencentes à superfamília *hAT* com uso do software MITE Tracker foi obtido um total de 1.277 sequências MITEs no genoma fêmea e 1.296 sequências MITEs no genoma macho com características de TEs *hAT*, as quais apresentaram tamanhos de 53 a 800 pb. Dentre estas sequências no genoma fêmea, 430 pertencem ao grupo Ac, 620 ao grupo Blackjack, 115 ao grupo Charlie e 112 ao grupo Tip100. Para o genoma macho foram obtidas 431 sequências do grupo Ac, 630 de Blackjack, 135 de Charlie e 100 de Tip100.

Em relação a análise das sequências TIRs destes elementos, as sequências TIRs representativas de cada elemento encontradas nesta análise foram utilizadas para agrupá-los nas quatro respectivas subfamílias (Tabela 5). Foram obtidas TIRs com tamanhos de 9 à 18 pb, onde muitas se apresentaram bem conservadas para grande parte destes elementos, com sequências idênticas ou quase idênticas observadas entre elementos diferentes (Tabela 5). Também foram demonstradas quais sequências apresentaram TSDs mais conservadas características de TSDs de TEs *hAT* adjacentes às TIRs (Tabela 5).

**Tabela 5:** Sequências de TIRs e TSDs obtidas para os elementos MITEs pertencentes ao grupo dos TEs *hAT* de *Apareiodon* sp. As sequências foram agrupadas de acordo com as quatro subfamílias mais abundantes observadas (Ac, Blackjack, Charlie e Tip100).

Sequência	TIR	TSD
<b>Ac</b>		
MITE_181	TAGGCCTGCCACGATAA	
MITE_77	TAGGCCTGCCGCGATAA	
MITE_57	TAGGCCTGCCGCGATA	
MITE_65	TAGGCCTGCCACGATA	GTTTTACT
MITE_107	TAGGGCTGCACGATAT	GTGGGTGG
MITE_430	TAGGGCTGCATGATAT	
MITE_178	TATAGGCCTGCT	
MITE_15	TCTAGGCCTGCC	
MITE_1478	TAGTGATGGGCA	
<b>Blackjack</b>		
MITE_1	TAGGGTGACCAGA	TGTA AAAAG
MITE_460	TAGGGTGACCATA	ATATAT
MITE_532	TAGGGTGACCATG	
MITE_1083	TAGGGTGACCAT	
MITE_2759	TAGGGTGACCATAT	GTAAAGCT
<b>Charlie</b>		
MITE_329	CAGGGGTCTCAACTGGT	
MITE_403	CAGGGGTCGGGAACCT	
MITE_2543	CAGTGGCGTGCACACA	
MITE_4802	CAGTGGTCTCAA ACT	CTTTAAAT
MITE_4799	CATGGTCTCAA ACT	CCTGAC
MITE_303	CACAGTTTGAGT	
<b>Tip100</b>		
MITE_4249	CAGTGGCGGTTCTAG	
MITE_4296	CAGTGGCGGTTCTGG	
MITE_4488	CAGGGGTGTAGTA	
MITE_4026	CAGGGGCGA	

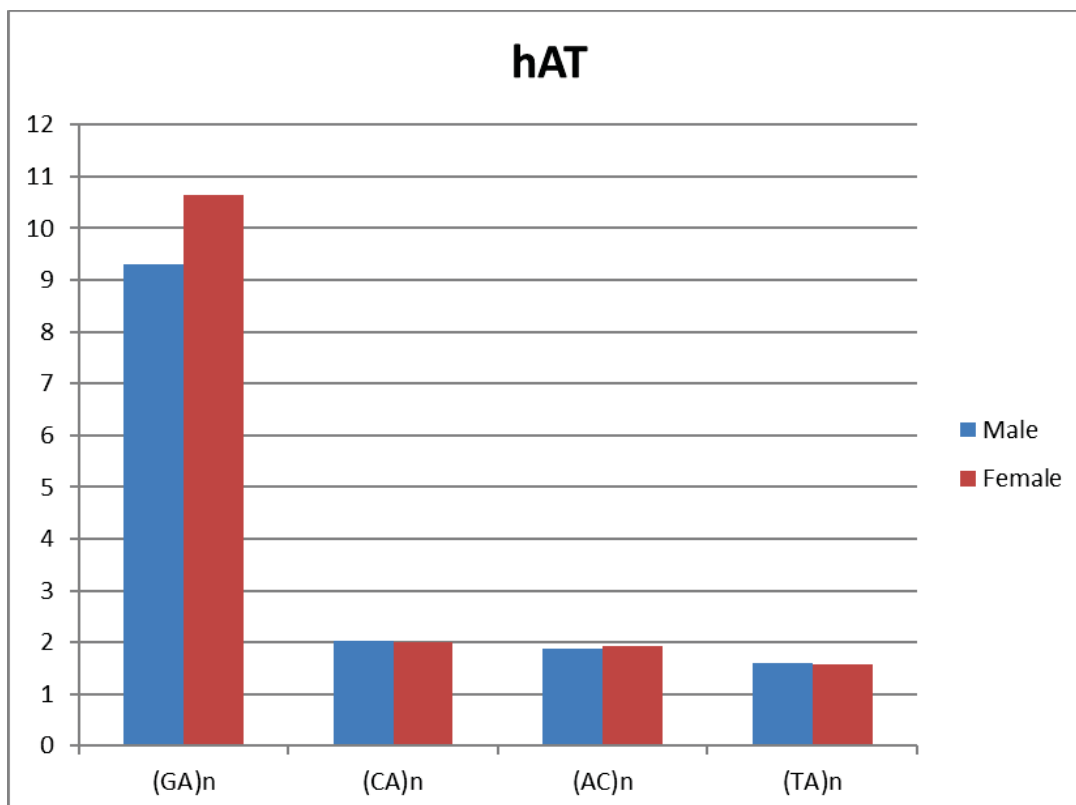
#### 4.6 Análise da presença de sequências repetidas *in tandem* associadas à *hAT*

Foi analisada a presença de sequências repetidas *in tandem* no interior das sequências *hAT*, onde foi possível verificar que algumas sequências do elemento *hAT* de *Apareiodon* sp. apresentaram associação com diferentes tipos de sequências microssatélites. Foram detectadas as sequências microssatélites (GA)<sub>n</sub>, (AC)<sub>n</sub>, (TAA)<sub>n</sub>, (CAT)<sub>n</sub>, (TA)<sub>n</sub>, (GAA)<sub>n</sub>, (GATA)<sub>n</sub>, (CA)<sub>n</sub>, (CAA)<sub>n</sub>, (CG)<sub>n</sub>, (CAGA)<sub>n</sub> e (CTAT)<sub>n</sub> internas à *hAT* em diferentes proporções entre elas no genoma *Apareiodon* sp. (Tabela 6).

**Tabela 6:** Estimativa da abundância (% dentro das sequências pertencentes a superfamília *hAT*) de sequências microssatélites encontradas nas bibliotecas *hAT* de macho e fêmea de *Apareiodon* sp.

<b>Microssatélite</b>	<b>Macho</b>	<b>Fêmea</b>
(GA)n	9,2918%	10,6522%
(AC)n	1,8784%	1,9397%
(TAA)n	0,0125%	0,0107%
(CAT)n	0%	0,0007%
(TA)n	1,6035%	1,5730%
(GAA)n	0,0011%	0,0005%
(GATA)n	0,0574%	0,0482%
(CA)n	2,0242%	1,9962%
(CAA)n	0,0003%	0,0003%
(CG)n	0,0001%	0%
(CAGA)n	0,0508%	0,0586%
(CTAT)n	0,0445%	0,0437%

Em relação a abundância destas sequências nos genomas, o microssatélite (GA)n foi o mais abundante no interior das sequências *hAT*, representando cerca de 9% da biblioteca *hAT* no genoma macho e 10% da biblioteca *hAT* em fêmea (Fig. 6) e contendo expansões com tamanho entre 10 e 388 pb. As sequências (AC)n, (TA)n e (CA)n apresentaram proporções em torno de 1-2% das sequências *hAT* em ambos os genomas (Fig. 6), e tamanhos das expansões variando entre 10 – 318, 10 – 292 e, 10 – 438 pb, respectivamente. Os demais microssatélites detectados [(TAA)n, (CAT)n, (GAA)n, (GATA)n, (CAA)n, (CG)n, (CAGA)n e (CTAT)n] apresentaram proporções menores que 1% nas bibliotecas *hAT* de ambos os genomas (Tabela 6).



**Figura 6:** Distribuição de DNAs microssatélite mais abundantes (> 1% dentro da biblioteca da superfamília *hAT*) pelas sequências dos TEs *hAT* de *Apareiodon* sp. dos genomas macho (azul) e fêmea (vermelho).

Contigs com a presença de sequências da superfamília *hAT* degeneradas no genoma fêmea foram submetidas a análises de suas sequências repetitivas através do software Dfam, com o objetivo de visualizar a distribuição de sequências repetitivas internas ou associadas à elementos *hAT*. Nesta análise foi possível de observar expansões de repetições microssatélites inseridas internamente ou entre fragmentos de sequências *hAT* degeneradas (Fig. 7). Ao mesmo tempo, também é possível a visualização do enriquecimento destes contigs com outras sequências repetitivas, onde foram observadas outras sequências de transposons de DNA e de retrotransposons, além de diferentes repetições de DNA microssatélite e satélite (Fig. 7).

Em relação aos diferentes tipos de sequências microssatélite observadas associadas a *hAT*, o NODE\_58801\_length\_2458\_cov\_9.53711 apresentou fragmentos de *hAT* associados com repetições TAT (Fig. 7). Já no NODE\_65599\_length\_1993\_cov\_16.1206 foram observadas repetições GT entre fragmentos *hAT* (Fig. 7). O NODE\_17394\_length\_13969\_cov\_9.47233 apresentou fragmentos menores similares a elementos *hAT* associados com repetições AC e TA (Fig. 7). O NODE\_2260\_length\_57153\_cov\_9.75052 demonstrou as repetições AC e AAGTCGCTAG

associadas a fragmentos *hAT* (Fig. 7). Já o NODE\_11682\_length\_20617\_cov\_16.1067 apresentou repetições A junto a fragmentos *hAT* (Fig. 7).

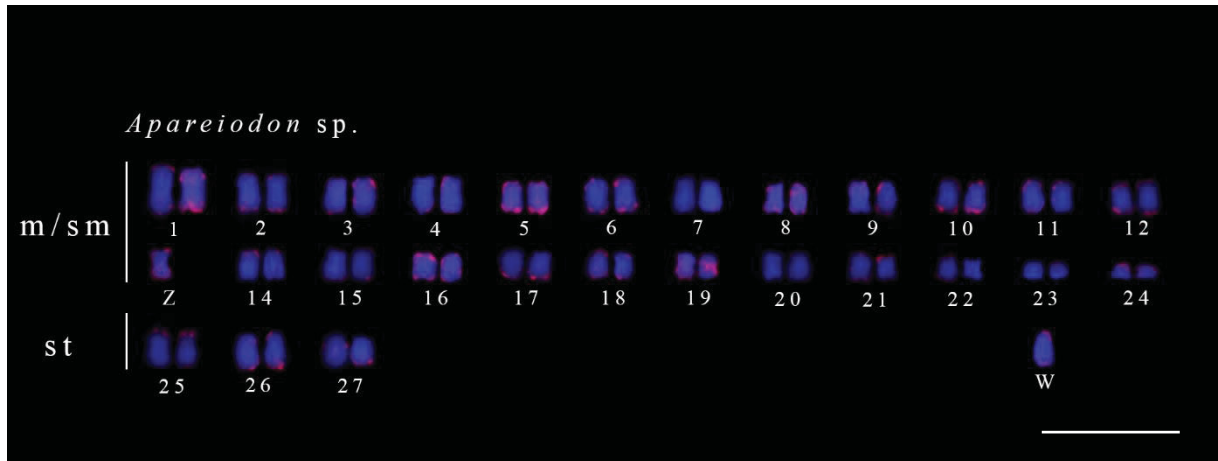


**Figura 7** – Cinco exemplos de contigs do genoma fêmea de *Apareiodon* sp. analisados através do software Dfam demonstrando a presença de sequências repetidas *in tandem* associadas a sequências *hAT*. São destacadas com \* e círculos em vermelho as sequências específicas de TEs do tipo *hAT* (laranja) com a presença de associações com sequências repetidas *in tandem* (preto). As identificações dos contigs são indicadas acima de cada imagem.

#### 4.7 Localização *in situ*

Cromossomos mitóticos da espécie *Apareiodon* sp. foram submetidos à técnica de FISH com a utilização de sondas obtidas com similaridade às sequências de elementos do tipo *hAT* desta espécie. A localização *in situ* demonstrou que elementos da superfamília *hAT* apresentam sua distribuição dispersa pelos cromossomos de *Apareiodon* sp. (Fig. 8).

Foram observados sinais da sonda *hAT* nas regiões terminais de quase todos os pares de cromossomos, com exceção dos pares 7, 20 e 23 (Fig. 8). Pequenos acúmulos em regiões intersticiais dos cromossomos foram observados principalmente no par 19, além de fracos sinais nos pares 1, 5, 8, além do cromossomo Z (Fig. 8). Em relação aos cromossomos sexuais Z e W, não foram detectados acúmulos diferenciais na região heteromórfica do cromossomo W (Fig. 8).



**Figura 8** – Cariótipo de *Apareiodon* sp. submetido à técnica de FISH com sonda da superfamília *hAT* (vermelho). Os cromossomos foram contracolorados com DAPI. Barra = 10  $\mu$ m.

## 5. DISCUSSÃO

### 5.1 Caracterização molecular dos TEs *hAT* no genoma de *Apareiodon* sp.

Os elementos das superfamílias *Tc1-Mariner* e *hAT* são os transposons de DNA mais abundantes no genoma de *Apareiodon* sp. (SCHEMBERGER et al., 2019). Por outro lado, poucos são os estudos sobre a caracterização molecular e a localização *in situ* de TEs *hAT* em Teleostei. Nesse grupo, dados existentes em relação a superfamília *hAT* focam no estudo do elemento *Tol2* em espécies marinhas (COSTA et al., 2013; FISCHER et al., 2004; HORI et al., 1998; KOGA et al., 2000, 2011) ou da diversidade e abundância de famílias *hAT* em *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus* e *Tetraodon nigroviridis* (GAO et al., 2016). Neste estudo, é apresentada uma primeira caracterização genômica e molecular de transposons de DNA *hAT* em uma espécie de peixe Neotropical detentora de sistema de cromossomos sexuais ZZ/ZW.

Elementos *hAT* com transposases funcionais apresentam tamanhos entre 2.000 à 5.000 pb (ATKINSON, 2014; ORTIZ; LORETO, 2009). Em *Apareiodon* sp., nas mais de 80.000 sequências *hAT* encontradas em ambos os genomas, apenas 92 sequências no genoma fêmea e 88 sequências no genoma macho apresentaram tamanhos maiores do que 1.000 pb e sem interrupções por sequências *in tandem* ou outros TEs. Da mesma forma, mais da metade destas sequências apresentaram tamanhos entre 80 - 200 pb e, domínios proteicos relacionados à transposase de *hAT* só foram encontrados em pouco mais de 2.000 sequências. De modo similar à todos TEs, elementos *hAT* apresentam a maior parte de suas cópias não autônomas ou degeneradas, e muitas vezes com a presença de poucas ou nenhuma cópia autônoma nos genomas eucarióticos (KARAKÜLAH; PAVLOPOULOU, 2018; RUBIN; LITHWICK; LEVY, 2001).

Elementos *hAT* autônomos foram descritos nas espécies dos peixes *Tetraodon nigroviridis* (FISCHER et al., 2004) e *Oryzias latipes* (HORI et al., 1998; KOGA et al., 2011; KOGA; HORI, 1999). Em comum, estes elementos *hAT* tiveram um tamanho entre 2,000 à 4,000 pb e TIRs de aproximadamente 20 pb (FISCHER et al., 2004; HORI et al., 1998). Em *Apareiodon* sp., apesar da detecção de vários domínios proteicos característicos de *hAT* com altos níveis de integridade de suas sequências, poucas sequências reuniram todas as condições para serem possivelmente autônomas. As transposases já conhecidas de *hAT* apresentam tamanho variando de 500 a 800 aminoácidos (LADEVÈZE et al., 2012; ROSSATO et al., 2014). ORFs de transposases *hAT* possivelmente íntegras no genoma de *Apareiodon* sp.

resultaram em tamanhos aproximados de 600 aminoácidos, entre as quais foi detectada pouca similaridade de resíduos de aminoácidos conservados em cada um dos quatro domínios proteicos. Estes resultados são condizentes com a provável ocorrência de diferentes famílias de *hAT* no genoma e corroboram as informações presentes na literatura, onde foi relatada a baixa similaridade das sequências dos aminoácidos nas sequências de diferentes tipos de transposases *hAT* (ATKINSON, 2014; ZHOU et al., 2004).

Ainda, nas análises das sequências de DNA que contém essas transposases possivelmente íntegras, em cinco sequências foram encontrados os elementos necessários à transcrição, bem como a observação de uma ORF transposase organizada em um único éxon, diferindo um pouco em relação ao encontrado para outras espécies de peixes, onde foram encontrados elementos com genes para transposase organizados em até quatro éxons (HORI et al., 1998; KOGA et al., 2011).

Na transposase *hAT*, a região N-terminal com domínios de dedos-de-zinco (ZF-BED) se caracteriza por sua função de ligação ao DNA, sendo assim responsável por ligar a transposase às TIRs e às regiões subterminais do TE (JIANG et al., 2016). Usualmente este domínio se apresenta relativamente conservado entre diferentes TEs (ORTIZ; LORETO, 2009). O segundo domínio de ligação ao DNA, localizado ao lado do domínio de dedos-de-zinco, também exerce a função de ligação entre a enzima transposase com as TIRs e regiões subterminais, além de possuir uma região importante para a dimerização da proteína (HICKMAN et al., 2005). Assim como descrito para outros organismos, esse segundo domínio de *hAT* e o seu domínio de inserção detectados no genoma de *Apareiodon* sp. representam as regiões menos conservadas da transposase e, desta forma, corrobora com a descrição que são as regiões mais variáveis entre elementos de diferentes famílias (ARENSBURGER et al., 2011).

O domínio catalítico DDE da transposase *hAT* está dividido em duas partes devido a presença do domínio de inserção entre elas (ATKINSON, 2014; YUAN; WESSLER, 2011). O domínio de inserção está localizado entre o segundo resíduo D e o resíduo E do domínio catalítico e é considerado a parte mais variável da transposase *hAT* (ATKINSON, 2014). Motivos conservados no interior do domínio de inserção são responsáveis por estabilizar a estrutura em grampo formada no início da transposição do TE (DAVIES et al., 2000; HICKMAN et al., 2005). Já o domínio catalítico é o responsável pela reação de clivagem das extremidades do TE e pela transferência do elemento para outra região do genoma (ATKINSON, 2014). Como característica, a segunda parte do domínio catalítico (Dimer tnp

hAT), localizada na porção C-terminal, a qual também é importante para a dimerização da proteína, mostra-se como sendo a região mais conservada entre as transposases quando comparada com vários elementos de diferentes famílias (ESSERS; ADOLPHS; KUNZE, 2000; KEMPKEN; WINDHOFER, 2001; WARREN; ATKINSON; O'BROCHTA, 1994).

A sequência (T/C)A(A/G)NG foi determinada como sendo um consenso presente nas TIRs do TE *hAT* (RUBIN; LITHWICK; LEVY, 2001). Como característica, as TIRs dos transposons de DNA *hAT* apresentam fracas sequências consenso, com tamanhos variando de 5 à 27 pb de baixa similaridade, mas que usualmente conservam uma adenina na segunda e uma guanina na quinta posição (ATKINSON, 2014). Ainda, os cinco primeiros pares de bases da TIR são os mais importantes do elemento, pois são os responsáveis por interagir com vários aminoácidos da transposase, em especial a adenina da segunda posição (ATKINSON, 2014). Na biblioteca *hAT* de *Apareiodon* sp. foram encontradas cinco sequências com ORFs possivelmente íntegras para transposase e com a presença de TIRs em suas extremidades, as quais apresentaram de 8 a 18 pb e conservaram os nucleotídeos ditos essenciais para a ligação da transposase. Com exceção destes nucleotídeos essenciais às TIRs, os demais segmentos exibiram sequências bem divergentes entre si nos TEs *hAT* de *Apareiodon*.

Os TEs *hAT* geram TSDs de 8 pb no momento de sua inserção, e assim como as TIRs, apresentam fracas sequências consenso, além de apresentarem características específicas para cada um dos grupos de *hAT* (ATKINSON, 2014). TEs do grupo Ac geram uma TSD com a sequência consenso 5' nTnnnnAn 3', enquanto o grupo Buster apresenta a sequência consenso 5' nnnTAnnn 3' (ARENSBURGER et al., 2011). Ainda não foi determinada uma sequência consenso para as TSDs do grupo Tip. Em *Apareiodon* sp. foram observadas três sequências que apresentaram TSDs adjacentes às TIRs, sendo apenas duas que apresentaram TSDs conservadas idênticas em ambas as porções 5' e 3'. Essas sequências que apresentaram TSDs conservadas detinham características de TSDs de elementos Buster, com as bases TA conservadas na quarta e quinta posição da TSD, sugerindo uma provável transposição recente destes elementos no genoma.

Assim sendo, é possível a inferência da presença de cinco sequências possivelmente autônomas no genoma de *Apareiodon* sp.: hAT-1\_AP, hAT-4\_AP, hAT-5\_AP, hAT-6\_AP e hAT-7AP, as quais apresentaram conservação de TIRs com as bases consideradas fundamentais para reações de ligação com domínios da transposase e uma ORF para a transposase com todos os domínios conservados. Dentre estas sequências, três ainda apresentam evidência de atividade recente devido a conservação de TSDs adjacentes às respectivas sequências de TIRs.

## 5.2 Agrupamentos e análise de diversidade e elementos *hAT* em *Apareiodon* sp.

Elementos da superfamília *hAT* são diversos e já foram organizados em três principais grupos e diversas famílias (ARENSBURGER et al., 2011; ROSSATO et al., 2014). O grupo Ac detém elementos amplamente distribuídos em plantas e animais e, agrupa diversas famílias já estudadas, como os primeiros elementos descritos Activator, Tam3 e Hobo (WARREN; ATKINSON; O'BROCHTA, 1994). Reúne também o elemento Hermes, o qual apresenta a estrutura molecular de sua transposase bem descrita (HICKMAN et al., 2005, 2018), além dos elementos Tol2 e Tgf2 descritos para peixes (HORI et al., 1998; JIANG et al., 2016; KOGA et al., 2000, 2011). O grupo Buster caracteriza-se por TEs predominantemente distribuídos entre os genomas animais, compreendendo os elementos das famílias Buster, a qual deu o nome ao grupo (ARENSBURGER et al., 2011), Charlie, inicialmente descobertos em humanos (SMIT, 1999), SPIN, descritos em diversas espécies animais (GILBERT et al., 2010; PACE et al., 2008), entre outras. Já o grupo Tip é o mais recentemente descrito, compreendendo elementos distribuídos entre plantas e animais, incluindo algumas famílias como Tip100, BuT2 de *Drosophila*, entre outras (ATKINSON, 2014; ROSSATO et al., 2014; ZHANG et al., 2013).

No presente trabalho os agrupamentos de sequências utilizando a transposase e o domínio Dimer tnp *hAT* para a análise da diversidade dos elementos da superfamília *hAT* demonstrou sequências representativas de quatro diferentes grupos (Ac, Blackjack, Charlie e Tip100) no genoma de *Apareiodon* sp., os quais compreendem TEs que se incluem nos três principais grandes grupos de elementos *hAT*: Ac (Ac), Buster (Charlie) e Tip (Tip100). Estes três principais grupos diferenciam-se principalmente por características específicas em suas transposases, especialmente entre o domínio de ligação ao DNA e o domínio de inserção, ou ainda pela estrutura das TSDs formadas no momento de sua inserção, havendo sequências específicas formadas para os diferentes grupos (ARENSBURGER et al., 2011). Dentro de cada grupo foi possível determinar ainda a presença de sequências similares a diferentes tipos de elementos, os quais corroboram para a diversidade de elementos *hAT* em *Apareiodon* sp.

Resultados semelhantes para a ocorrência das famílias *hAT* presentes em *Apareiodon* sp. também foram descritos nos genomas de outros peixes, como *Paralichthys olivaceus*, *Solea senegalensis*, *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus* e *Tetraodon nigroviridis* (GAO et al., 2016; GARCÍA et al., 2019; NAM et al., 2018). Em comparação aos dados genômicos dos outros peixes, onde poucas ou nenhuma sequência da família Blackjack foi identificada (GAO et al., 2016; GARCÍA et al., 2019; NAM et al., 2018), essa família

apresentou um número de cópias significativo em *Apareiodon* sp. Contudo, poucas foram as sequências desta família a apresentar transposase ou domínios proteicos íntegros, resultados corroborados com a análise das árvores ML, que também demonstraram a proximidade destes elementos àqueles do grupo Ac. Os resultados dos agrupamentos demonstram que a ampla maioria de cópias *hAT* no genoma *Apareiodon* sp. são de elementos do grupo Ac, onde, pela análise da árvore com sequências do domínio C-terminal da transposase, é possível perceber a distribuição de sequências Ac em uma maior diversidade de famílias. Já pela análise da árvore com as sequências completas da transposase é possível a identificação de uma maior abundância de sequências mais conservadas dentro da família Charlie, na qual estiveram presentes as cinco sequências previamente descritas como aquelas que apresentaram todas as características necessárias para serem elementos autônomos no genoma (hAT-1\_AP, hAT-4\_AP, hAT-5\_AP, hAT-6\_AP e hAT-7AP), inferindo na maior conservação de elementos hAT-Charlie no genoma de *Apareiodon* sp. e na possibilidade de esta família ainda estar ativa no genoma. Consequentemente, os resultados obtidos reforçam para a diversidade desta superfamília de transposons de DNA descrita em Teleostei.

### 5.3 Estimativa do tempo de inserção dos elementos *hAT*

As composições de TEs nos genomas de peixes são bem divergentes, onde os TEs apresentam padrões de inserção nos genomas e níveis de substituição nucleotídica muito diferentes entre espécies (CHALOPIN et al., 2015b; SHAO et al., 2019). Em uma análise da composição total de TEs, Schemberger et al. (2019) sugeriram duas ondas principais de invasão de TEs no genoma de *Apareiodon*. No presente estudo, o detalhamento da inserção de elementos de diferentes famílias *hAT* demonstrou que sequências Tip100 se integraram no genoma de *Apareiodon* em poucas cópias em um período estimado de 15 milhões de anos atrás. Já os elementos Ac, Blackjack e Charlie também apresentaram inserção de cópias no período de 15 milhões de anos atrás, porém uma maior amplificação genômica destes elementos é estimada a ocorrer a partir de 5 milhões de anos. De modo similar, uma fase de invasão relativamente recente de TEs *hAT* também foi observada para a espécie *Oryzias latipes*, onde foram encontradas cópias conservadas do elemento *Tol2* (KOGA et al., 2000, 2011; KOGA; HORI, 1999). Gao et al. (2016) analisaram a distribuição no genoma e os níveis de substituição nucleotídica dos TEs *hAT* em quatro espécies de peixes, observando que nestes genomas, estes elementos apresentam baixos níveis de substituição nucleotídica, além de apresentarem uma grande diversidade de famílias *hAT*. Embora estas espécies analisadas por Gao et al. (2016) não

sejam filogeneticamente próximas à *Apareiodon* sp., os dados comparados permitem identificar uma grande diversidade de famílias *hAT* e um viés de inserção relativamente recente deste transposon de DNA para alguns genomas de peixes.

Aliado a estimativa de tempo de invasão de elementos *hAT* em *Apareiodon* sp., a observação de poucas cópias com sequências putativamente autônomas reflete que a grande maioria destes TEs estão silenciados no genoma e, atualmente passam pelo estágio de deterioração molecular de sequências. De acordo com essa proposta, os elementos *hAT* de *Apareiodon* sp. estão no início da chamada fase de senescência descrita por Kidwell e Lisch (2001), onde não ocorre mais amplificação das cópias destes TEs e poucos são os possíveis elementos autônomos no genoma. Assim, os elementos acumulam mutações em suas sequências, tornando-os inativos e neutros ou, em alguns casos, sendo cooptados para outras funções no genoma (MILLER; McDONALD e PINSKER, 1997). No entanto, ainda é importante ressaltar a presença de cópias *hAT*-Charlie possivelmente ativas no genoma, assim, apesar da grande maioria das cópias representarem elementos degenerados, estas cópias potencialmente ativas ainda apresentam a possibilidade de contribuir com novas inserções e modificações no genoma.

#### 5.4 Análise das sequências MITEs pertencentes ao grupo dos TEs *hAT*

MITEs pertencentes aos quatro grupos *hAT* visualizadas em *Apareiodon* sp. foram identificados no genoma. A maior representatividade de sequências MITEs foi para elementos Blackjack, seguido por elementos Ac. Esses resultados demonstram um maior viés de formação de MITEs na família Blackjack, a qual, na análise dos agrupamentos havia apresentado um menor número de cópias totais quando comparada com outros grupos, dessa forma, corroborando a descrição do menor número de cópias Blackjack apresentando domínios para a transposase. Poucos são os estudos envolvendo a descrição de MITEs *hAT* em genomas de peixes, mas estes também demonstraram a maior presença de MITEs pertencentes ao grupo Ac, como os descritos no genoma da espécie *Oryzias latipes* (KOGA et al., 2011).

A alta conservação das TIRs encontradas nos MITEs das famílias *hAT* de *Apareiodon* sp. indicam que, mesmo sem a autonomia característica destas sequências, essas poderiam apresentar transposição no genoma a partir do uso de transposases de elementos autônomos, como amplamente proposto em outros estudos (FATTASH et al., 2013; JIANG et al., 2004; SHIRASAWA et al., 2012). Já a presença de poucos elementos com TSDs características de *hAT* conservadas infere em uma invasão menos recente para estes elementos. Os dados relativos

à transposase e estrutura molecular do TE demonstraram elementos pertencentes à família Charlie putativamente autônomos no genoma de *Apareiodon*. No entanto, pequena foi a representatividade de MITEs da subfamília Charlie, embora a alta conservação de suas TIRs possa representar inserção recente no genoma. De modo similar, a alta quantidade de MITEs Blackjack e Ac pode representar atividade transposase recente para esta família de TEs *hAT*. Ainda, com a comparação destas TIRs obtidas para MITEs com as TIRs das sequências íntegras foi observado que, apesar de não terem sido obtidas sequências TIRs de MITEs idênticas àquelas dos elementos íntegros, as sequências de TIRs da família Charlie dos MITEs apresentaram similaridade a algumas TIRs dos elementos íntegros, em especial dos TEs *hAT-1\_AP*, *hAT-6\_AP* e *hAT-7\_AP*, inferindo na possibilidade de estas sequências estarem envolvidas na mobilização de MITEs dentro do genoma.

Além da mobilidade de sequências MITEs, estes elementos são descritos ter influência em relação a evolução do genoma, principalmente na geração de novidades evolutivas atreladas a eventos de controle da expressão gênica (FESCHOTTE, 2008; FESCHOTTE; PRITHAM, 2007). Estudos de MITEs pertencentes à *hAT* já observaram essas sequências localizadas em regiões ricas em genes gerando sítios regulatórios, bem como, já foram descritos originar sequências de microRNAs e lncRNAs (MENZEL et al., 2014; PERUMAL et al., 2020; WANG et al., 2016). Dessa forma, a anotação genômica destas sequências pode abrir cenário de investigação para coopção de MITEs derivados de *hAT* para controle da expressão de genes em *Apareiodon*.

### 5.5 Associação de sequências repetidas *in tandem* à elementos *hAT* degenerados

A associação entre TEs senescentes e a origem ou dispersão de sequências satélites, mini- e microssatélites é amplamente proposta na literatura (KIDWELL, 2002; PAÇO; FREITAS; VIEIRA-DA-SILVA, 2019; WILDER; HOLLOCHER, 2001). É conhecido que algumas sequências de repetição *in tandem* tiveram origem a partir de TEs ou, no caso dos DNAs satélites, apresentam alguns segmentos diferenciados a partir de TEs (KIDWELL, 2002; MEŠTROVIĆ et al., 2015; PAÇO; FREITAS; VIEIRA-DA-SILVA, 2019). Foi proposto também que repetições *in tandem* são mais frequentemente encontradas em transposons de DNA do que em retrotransposons (BISCOTTI; OLMO; HESLOP-HARRISON, 2015), onde as sequências das TIRs parecem ter maior relação com a origem das cópias de DNAs satélites (MEŠTROVIĆ et al., 2015). Sequências microssatélites podem se originar e se multiplicar no interior de TEs a partir de recombinação ectópica ou pela amplificação de sequências internas

pré-existentes (CHISTIYAKOV; HELLEMANS; VOLCKAERT, 2006; MEŠTROVIĆ et al., 2015). Em outra via, durante a duplicação do DNA, a DNA polimerase “escorrega” durante a religação do *clamp* ao passar por uma região com repetições curtas, e assim, reinicia a síntese da região já duplicada (LEVINSON; GUTMAN, 1987). A partir deste modelo, sequências de TEs contendo proto-microsatélites podem dar origem a repetições simples, levando a expansão destas unidades nos genomas (LEVINSON; GUTMAN, 1987; WILDER; HOLLOCHER, 2001).

Poucos são os estudos que demonstraram origem/colocalização de sequências *in tandem* com elementos da superfamília *hAT*. Em *Tetraodon nigroviridis* foi observada a colocalização de sequências minissatélite com o elemento *Tol2*, as quais apresentam acúmulo em regiões heterocromáticas (FISCHER et al., 2004). Os dados obtidos em *Apareiodon* sp. demonstraram que diferentes tipos de sequências microsatélites foram detectadas no interior de sequências *hAT* degeneradas, sendo a sequência (GA)<sub>n</sub> a mais abundante. Dados demonstraram que o microsatélite (GA)<sub>n</sub> foi o terceiro mais abundante no genoma de *Apareiodon* sp. e inúmeras vezes colocalizados com regiões ricas em TEs (SCHEMBERGER et al., 2019). Desse modo, é provável a participação de cópias degeneradas do TE *hAT* para origem de microsatélites, as quais podem, por meio de seus mecanismos de expansão pelo genoma, contribuir para a evolução do genoma e variação cromossômica em Parodontidae.

A análise dos contigs que contém elementos *hAT* associados com sequências *in tandem* demonstrou que as expansões microsatélites podem, por vezes, interromper a similaridade das sequências *hAT* ao acumular em determinadas regiões, dividindo um TE *hAT* deteriorado em pequenos segmentos. Ainda, em vários contigs foi observada a presença de sequências *hAT* degeneradas em regiões com proximidade a outras expansões *in tandem* e também a fragmentos de diversos outros TEs, demonstrando que a presença destes TEs *hAT* degenerados associados a sequências repetidas *in tandem* provavelmente ocorre predominantemente em regiões de ricas em DNAs repetitivos.

#### 5.6 Localização *in situ* de TEs *hAT* nos cromossomos de *Apareiodon* sp.

A espécie *Apareiodon* sp. apresenta  $2n = 54$  e sistema de cromossomos sexuais do tipo ZZ/ZW (VICARI et al., 2006). A diferenciação da região heteromórfica do cromossomo W de *Apareiodon* sp. foi proposta ocorrer a partir da invasão dos transposons de DNA *Helitron*, *En/Spm* e *Tc1-Mariner* iniciada a cerca de 20 milhões de anos atrás (SCHEMBERGER et al., 2019). As estimativas de tempo de inserção de sequências da subfamília *hAT* no genoma de

*Apareiodon* sp. demonstraram pequeno acúmulo destes elementos a cerca de 15 milhões de anos atrás, seguido por uma invasão de um grande número de cópias a partir de 5 milhões de anos atrás. Contudo, embora a abundância de cópias *hAT* tenha sido maior no genoma fêmea em comparação ao macho, nenhum sítio de grande acúmulo foi visualizado na região heteromórfica do cromossomo W. Dessa forma, é provável que sequências *hAT* tenham se inserido de modo disperso nos cromossomos sexuais de *Apareiodon* sp., principalmente em suas regiões terminais.

A localização de acúmulos *hAT* preferencialmente em regiões terminais dos cromossomos foi também descrita para o mapeamento in situ de sequências (GA)<sub>n</sub> (SCHEMBERGER et al., 2019). Esses dados corroboram com a proposta de deterioração de *hAT* e origem deste microssatélite em regiões heterocromáticas terminais. Por outro lado, Schemberger et al. (2019) também relataram um acúmulo (GA)<sub>n</sub> na região heteromórfica do cromossomo W, o qual, com os dados aqui relatados, não apresenta relação com sítios *hAT*. Desse modo, a localização dispersa de sequências *hAT* nos cromossomos de *Apareiodon* sp. parece ter relação com sua invasão relativamente recente no genoma.

## 6. CONCLUSÃO

Neste estudo, sequências de transposons de DNA *hAT* foram caracterizadas no genoma de uma espécie de peixe Neotropical da família Parodontidae. Os resultados demonstraram que TEs *hAT* apresentam um grande número de cópias inseridas no genoma de *Apareiodon* sp., as quais se encontram distribuídas predominantemente em quatro grupos diferentes (Ac, Blackjack, Charlie e Tip100), sendo as sequências Ac as mais abundantes. A maior parte destas sequências apresentam um período de inserção no genoma relativamente recente, com exceção dos TEs da família Tip100, que apresentaram um menor número de sequências e um período de invasão mais antigo em relação às outras famílias.

A maior parte das sequências *hAT* no genoma desta espécie apresentam-se degeneradas, porém foram identificadas cinco sequências pertencentes à família Charlie que reuniram todas as características necessárias para serem elementos íntegros e autônomos, conseqüentemente ainda tendo a possibilidade de estarem ativas e gerarem modificações no genoma. Os dados, portanto, demonstraram que as cópias do TE *hAT* estão no início da fase de senescência no genoma de *Apareiodon* sp., onde poucas sequências reuniram todos os elementos para serem putativamente autônomas. Ainda, sequências MITEs correspondentes aos quatro grupos *hAT* analisadas foram identificadas no genoma, com maior abundância de sequências Blackjack. Aliada à provável pequena ou falta de mobilização do TE *hAT* em *Apareiodon* sp., o estudo demonstrou que uma parte das suas sequências degeneradas podem estar envolvidas na origem de repetições microssatélites, com base na observação de diversas sequências microssatélites associadas a *hAT*. Por fim, a localização das sequências *hAT* nos cromossomos de *Apareiodon* sp. demonstrou distribuição dispersa destes elementos pelo cariótipo e sem relação com a diferenciação da região heteromórfica do cromossomo W.

Deste modo, este estudo além de apresentar uma primeira caracterização molecular para elementos *hAT* em um genoma de peixe Neotropical, abre perspectiva de avaliação de sequências que podem ter sido cooptadas em *Apareiodon* sp., principalmente àquelas relacionadas ao controle da regulação gênica.

## REFERÊNCIAS

- ARENSBURGER, P. et al. Phylogenetic and functional characterization of the hAT transposon superfamily. **Genetics**, v. 188, n. 1, p. 45–57, 2011.
- ARKHIPOVA, I.; MESELSON, M. Deleterious transposable elements and the extinction of asexuals. **BioEssays**, v. 27, n. 1, p. 76–85, 2005.
- ARKHIPOVA, I. R. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. **Mobile DNA**, v. 8, n. 1, p. 1–14, 2017.
- ATKINSON, P. W. hAT Transposable Elements. **Microbiology Spectrum**, v. 2, n. 3, p. 1–27, 2014.
- BENSON, G. Tandem repeats finder: A program to analyze DNA sequences. **Nucleic Acids Research**, v. 27, n. 2, p. 573–580, 1999.
- BERTOLLO, L.A.C., CIOFFI, M.B., MOREIRA-FILHO, O. Direct Chromosome preparations from freshwater teleost fishes. In: Ozouf-Costaz, C., Pisano, E., Foresti, F., Toledo, L. F. de A. (Eds). Fish techniques, Ray-Fin Fishes and Chondrichthyans, **CRC Press**, p. 21-26, 2015.
- BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. (PAT. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415–420, 2015.
- BOURQUE, G. et al. Ten things you should know about transposable elements. **Genome biology**, v. 19, n. 1, p. 199, 2018.
- 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, 2015a.
- CHALOPIN, D. et al. Transposable elements and early evolution of sex chromosomes in fish. **Chromosome Research**, v. 23, n. 3, p. 1–16, 2015b.
- CHARLESWORTH, B. Evolution of Sex Chromosomes. **Science**, v. 251, p. 1030–1033, 1991.
- CHARLESWORTH, D.; CHARLESWORTH, B.; MARAIS, G. Steps in the evolution of heteromorphic sex chromosomes. **Heredity**, v. 95, n. 2, p. 118–128, 2005.
- CHÉNAIS, B. et al. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. **Gene**, v. 509, n. 1, p. 7–15, 2012.
- CHISTIYAKOV, D. A.; HELLEMANS, B.; VOLCKAERT, F. A. M. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. **Aquaculture**, v. 255, n. 1–4, p. 1–29, 2006.
- COSTA, G. W. W. F. et al. Transposable Elements in Fish Chromosomes : A Study in the Marine Cobia Species. **Cytogenetic and Genome Research**, p. 1–7, 2013.
- CRESCENTE, J. M. et al. MITE Tracker: An accurate approach to identify miniature inverted-repeat transposable elements in large genomes. **BMC Bioinformatics**, v. 19, n. 1, p. 1–10, 2018.
- DAVIES, D. R. et al. Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate. **Science**, v. 289, n. 5476, p. 77–85, 2000.
- DHILLON, B. et al. Variable genome evolution in fungi after transposon-mediated amplification of a housekeeping gene. **Mobile DNA**, v. 10, n. 1, p. 1–16, 2019.

- DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin**, v. 19, p. 11–15, 1990.
- EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.
- EICHLER, E. E.; SANKOFF, D. Structural Dynamics of Eukaryotic Chromosome Evolution. **Science**, v. 301, p. 793–797, 2003.
- EICKBUSH, T. H.; JAMBURUTHUGODA, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. **Virus Research**, v. 134, n. 1–2, p. 221–234, 2008.
- EL-GEBALI, S. et al. The Pfam protein families database in 2019. **Nucleic Acids Research**, v. 47, n. D1, p. D427–D432, 2019.
- ELLISON, C. E.; BACHTROG, D. Dosage compensation via transposable element mediated rewiring of a regulatory network. **Science**, v. 342, n. 6160, p. 846–850, 2013.
- ESSERS, L.; ADOLPHS, R. H.; KUNZE, R. A highly conserved domain of the maize Activator transposase is involved in dimerization. **Plant Cell**, v. 12, n. 2, p. 211–223, 2000.
- FATTASH, I. et al. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. **Genome**, v. 56, p. 475–486, 2013.
- FERNÁNDEZ-MEDINA, R. D. et al. Losing identity: structural diversity of transposable elements belonging to different classes in the genome of *Anopheles gambiae*. **BMC Genomics**, v. 13, n. 1, p. 272, 2012.
- FESCHOTTE, C. The contribution of transposable elements to the evolution of regulatory networks. **Nature Reviews Genetics**, v. 9, n. 5, p. 397–405, 2008.
- FESCHOTTE, C.; PRITHAM, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. **Annu Rev Genet**, v. 41, p. 331–368, 2007.
- FINNEGAN, D. J. Eukaryotic Transposable Elements and Genome Evolution. **Trends Genet**, v. 5, n. 4, p. 103–107, 1989.
- FISCHER, C. et al. Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish *Tetraodon nigroviridis*. **Gene**, v. 336, n. 2, p. 175–183, 2004.
- GAO, B. et al. The contribution of transposable elements to size variations between four teleost genomes. **Mobile DNA**, v. 7, n. 1, p. 1–16, 2016.
- GAO, B. et al. Evolution and Domestication of Tc1/mariner transposons in the African coelacanth (*Latimeria chalumnae*) genome. **Genome**, v. 1, n. 12, 2020.
- GARCÍA, E. et al. Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a Rex transposon located in a proto-sex chromosome. **Scientific Reports**, v. 9, n. 17146, p. 1–14, 2019.
- GILBERT, C. et al. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. **Nature**, v. 464, n. 7293, p. 1347–1350, 2010.
- GLUGOSKI, L. et al. Co-located hAT transposable element and 5S rDNA in an interstitial telomeric sequence suggest the formation of Robertsonian fusion in armored catfish. **Gene**, v. 650, p. 49–54, 2018.
- GRABUNDZIJA, I. et al. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. **Nature Communications**, v. 7, n. 10716, p. 1–

12, 2016.

GUO, C. et al. Genome-Wide Comparative Analysis of Miniature Inverted Repeat Transposable Elements in 19 *Arabidopsis thaliana* Ecotype Accessions. **Scientific Reports**, v. 7, n. 1, p. 1–12, 2017.

HAVECKER, E. R.; GAO, X.; VOYTAS, D. F. The diversity of LTR retrotransposons. **Genome Biology**, v. 5, n. 6, 2004.

HERPIN, A. et al. Transcriptional rewiring of the sex determining *dmrt1* gene duplicate by transposable elements. **PLoS Genetics**, v. 6, n. 2, p. 1–15, 2010.

HICKMAN, A. B. et al. Molecular architecture of a eukaryotic DNA transposase. **Nature Structural and Molecular Biology**, v. 12, n. 8, p. 715–721, 2005.

HICKMAN, A. B. et al. Structural insights into the mechanism of double strand break formation by HERMES, a hAT family eukaryotic DNA transposase. **Nucleic Acids Research**, v. 46, n. 19, p. 10286–10301, 2018.

HORI, H. et al. An active Ac-like transposable element in teleost fish. **Journal of Marine Biotechnology**, v. 6, n. 4, p. 206–207, 1998.

JANGAM, D.; FESCHOTTE, C.; BETRÁN, E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. **Trends in Genetics**, v. 33, n. 11, p. 817–831, 2017.

JIANG, N. et al. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). **Current Opinion in Plant Biology**, v. 7, n. 2, p. 115–119, 2004.

JIANG, X. Y. et al. The N-terminal zinc finger domain of Tgf2 transposase contributes to DNA binding and to transposition activity. **Scientific Reports**, v. 6, n. May, p. 1–12, 2016.

JORDAN, I. K. et al. Origin of a substantial fraction of human regulatory sequences from transposable elements. **Trends in Genetics**, v. 19, n. 2, p. 68–72, 2003.

KAPITONOV, V. V.; JURKA, J. Rolling-circle transposons in eukaryotes. **Proceedings of the National Academy of Sciences**, v. 98, n. 15, p. 8714–8719, 2001.

KAPITONOV, V. V.; JURKA, J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. **PLoS Biology**, v. 3, n. 6, p. 0998–1011, 2005.

KAPITONOV, V. V.; JURKA, J. Helitrons on a roll: eukaryotic rolling-circle transposons. **Trends in Genetics**, v. 23, n. 10, p. 521–529, 2007.

KARAKÜLAH, G.; PAVLOPOULOU, A. In silico phylogenetic analysis of hat transposable elements in plants. **Genes**, v. 9, n. 284, p. 1–13, 2018.

KEMPKEN, F.; WINDHOFER, F. The hAT family: A versatile transposon group common to plants, fungi, animals, and man. **Chromosoma**, v. 110, n. 1, p. 1–9, 2001.

KIDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes. **Genetica**, v. 115, p. 49–63, 2002.

KIDWELL, M. G.; LISCH, D. R. Perspective: Transposable Elements, Parasitic DNA, and Genome Evolution. **Evolution**, v. 55, n. 1, p. 1–24, 2001.

KOGA, A. et al. Evidence for recent invasion of the medaka fish genome by the Tol2 transposable element. **Genetics**, v. 155, n. 1, p. 273–281, 2000.

KOGA, A. et al. Occurrence of a short variant of the Tol2 transposable element in natural populations of the medaka fish. **Genetics Research**, v. 93, n. 1, p. 13–21, 2011.

- KOGA, A.; HORI, H. Homogeneity in the structure of the medaka fish transposable element Tol2. **Genetical Research**, v. 73, n. 1, p. 7–14, 1999.
- KOHANY, O. et al. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **BMC Bioinformatics**, v. 7, p. 1–7, 2006.
- KOJIMA, K. K. Structural and sequence diversity of eukaryotic transposable elements. **Genes & Genetic Systems**, p. 1–20, 2018.
- KUMAR, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 2018.
- LADEVÈZE, V. et al. General survey of hAT transposon superfamily with highlight on hobo element in Drosophila. **Genetica**, v. 140, n. 7–9, p. 375–392, 2012.
- LAVIALLE, C. et al. Paleovirology of ‘ syncytins ’, retroviral env genes exapted for a role in placentation. **Phil Trans R Soc B**, v. 368, n. 20120507, p. 1–10, 2013.
- LEVAN, A.; FREDGA, K.; SANDBERG, A. A. Nomenclature for centromeric position on chromosomes. **Hereditas**, v. 52, p. 201–220, 1964.
- LEVINSON, G.; GUTMAN, G. A. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. **Molecular Biology and Evolution**, v. 4, n. 3, p. 203–221, 1987.
- LI, S. F. et al. The landscape of transposable elements and satellite DNAs in the genome of a dioecious plant spinach (*Spinacia oleracea* L.). **Mobile DNA**, v. 10, n. 1, p. 1–15, 2019.
- LONGO, M. S. et al. Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. **BMC Genomics**, v. 10, p. 1–14, 2009.
- MCCLINTOCK, B. Cytogenetic studies of maize and Neurospora. **Carnegie Inst. Washington**, v. 46, p. 146–152, 1947.
- MENZEL, G. et al. The diversification and activity of hAT transposons in Musa genomes. **Chromosome Research**, v. 22, n. 4, p. 559–571, 2014.
- MEŠTROVIĆ, N. et al. Structural and functional liaisons between transposable elements and satellite DNAs. **Chromosome Research**, v. 23, n. 3, p. 583–596, 2015.
- MILLER, W. J.; MCDONALD, J. F.; PINSKER, W. Molecular domestication of mobile elements. **Genetica**, v. 100, p. 261–270, 1997.
- NAM, G. H. et al. Identification of transposable elements fused in the exonic region of the olive flounder genome. **Genes and Genomics**, v. 40, n. 7, p. 707–713, 2018.
- ORTIZ, M. DE F.; LORETO, E. L. S. Characterization of new hAT transposable elements in 12 Drosophila genomes. **Genetica**, v. 135, n. 1, p. 67–75, 2009.
- OSTERTAG, E. M.; KAZAZIAN, H. H. Biology of Mammalian L1 Retrotransposons. **Annu Rev Genet**, v. 35, n. 1, p. 501–538, 2001.
- PACE, J. K. et al. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 44, p. 17023–17028, 2008.
- 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, n. 12, 2019.
- PERUMAL, S. et al. Characterization of B-Genome Specific High Copy hAT MITE Families in Brassica nigra Genome. **Frontiers in Plant Science**, v. 11, n. July, p. 1–12, 2020.

- PEVZNER, P.; TESLER, G. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. **Genome research**, v. 13, n. 1, p. 37–45, 2003.
- PINKEL, D.; STRAUME, T.; GRAY, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 9, p. 2934–2938, 1986.
- PLATT, R. N.; BLANCO-BERDUGO, L.; RAY, D. A. Accurate transposable element annotation is vital when analyzing new genome assemblies. **Genome Biology and Evolution**, v. 8, n. 2, p. 403–410, 2016.
- ROSSATO, D. O. et al. BuT2 is a member of the third major group of hAT transposons and is involved in horizontal transfer events in the genus *Drosophila*. **Genome Biology and Evolution**, v. 6, n. 2, p. 352–365, 2014.
- ROZEN, S., SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers. **Methods in Molecular Biology**, 132, 365–86, 2000.
- RUBIN, E.; LITHWICK, G.; LEVY, A. A. Structure and evolution of the hAT transposon superfamily. **Genetics**, v. 158, n. 3, p. 949–957, 2001.
- SCHARTL, M. et al. Sox5 is involved in germ-cell regulation and sex determination in medaka following co-option of nested transposable elements. **BMC Biology**, v. 16, n. 1, p. 1–17, 2018.
- SCHEMBERGER, M. O. et al. Differentiation of repetitive DNA sites and sex chromosome systems reveal closely related group in Parodontidae (Actinopterygii: Characiformes). **Genetica**, v. 139, n. 11–12, p. 1499–1508, 2011.
- SCHEMBERGER, M. O. et al. Construction and Characterization of a Repetitive DNA Library in Parodontidae (Actinopterygii: Characiformes): A Genomic and Evolutionary Approach to the Degeneration of the W Sex Chromosome. **Zebrafish**, v. 11, n. 6, p. 518–527, 2014.
- SCHEMBERGER, M. O. et al. Sequence analyses and chromosomal distribution of the Tc1/Mariner element in Parodontidae fish (Teleostei: Characiformes). **Gene**, v. 593, n. 2, p. 308–314, 2016.
- SCHEMBERGER, M. O. et al. DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. **Chromosoma**, 2019.
- SHAO, F. et al. FishTEDB: A collective database of transposable elements identified in the complete genomes of fish. **Database**, v. 2018, p. 1–9, 2018.
- SHIRASAWA, K. et al. Characterization of active miniature inverted-repeat transposable elements in the peanut genome. **Theoretical and Applied Genetics**, v. 124, n. 8, p. 1429–1438, 2012.
- ŚLIWIŃSKA, E. B.; MARTYKA, R.; TRYJANOWSKI, P. Evolutionary interaction between W/Y chromosome and transposable elements. **Genetica**, v. 144, n. 3, p. 267–278, 2016.
- SMIT, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. **Current Opinion in Genetics and Development**, v. 9, n. 6, p. 657–663, 1999.
- SMIT, A. F. A.; HUBLEY, R.; GREEN, P. RepeatMasker Open–4.0, 2013–2015. Disponível em: <http://www.repeatmasker.org>.
- SOBREIRA, T. J. P.; DURHAM, A. M.; GRUBER, A. TRAP: Automated classification, quantification and annotation of tandemly repeated sequences. **Bioinformatics**, v. 22, n. 3, p. 361–362, 2006.

- SOLOVYEV, V. et al. Automatic annotation of eukaryotic genes, pseudogenes and promoters. **Genome biology**, v. 7 Suppl 1, n. Suppl 1, p. 1–12, 2006.
- SOTERO-CAIO, C. G. et al. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. **Genome Biology and Evolution**, v. 9, n. 1, p. 161–177, 2017.
- STEINEMANN, M.; STEINEMANN, S. Enigma of Y chromosome degeneration: Neo-Y and Neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. **Genetica**, v. 102, n. 103, p. 409–420, 1998.
- STORER, J. et al. The Dfam community resource of transposable element families, sequence models, and genome annotations. **Mobile DNA**, v. 12, n. 1, p. 1–14, 2021.
- THOMAS, J.; PRITHAM, E. J. Helitrons, the Eukaryotic Rolling-circle Transposable Elements. **Microbiology Spectrum**, v. 3, n. 4, p. 1–32, 2015.
- TRALDI, J. B. et al. Chromosome Mapping of H1 and H4 Histones in Parodontidae (Actinopterygii: Characiformes): Dispersed and / or Co-Opted Transposable Elements? **Cytogenetic and Genome Research**, p. 1–8, 2019.
- UNTERGASSER, A.; NIJVEEN, H. Primer 3Plus. 2007. Disponível em: <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>.
- VAN DE LAGEMAAT, L. N. et al. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. **Trends in Genetics**, v. 19, n. 10, p. 530–536, 2003.
- VICARI, M. R. et al. ZZ/ZW sex chromosome system in an undescribed species of the genus *Apareiodon* (Characiformes, Parodontidae). **Cytogenetic and Genome Research**, v. 114, n. 2, p. 163–168, 2006.
- WANG, L. et al. Evolutionary origin of Rosaceae-specific active non-autonomous hAT elements and their contribution to gene regulation and genomic structural variation. **Plant Molecular Biology**, v. 91, n. 1–2, p. 179–191, 2016.
- WARREN, I. A. et al. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. **Chromosome Research**, v. 23, n. 3, p. 505–531, 2015.
- WARREN, W. D.; ATKINSON, P. W.; O'BROCHTA, D. A. The Hermes transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the hobo, Ac, and Tam3 (hAT) element family. **Genetical Research**, v. 64, n. 2, p. 87–97, 1994.
- WERREN, J. H. Selfish genetic elements, genetic conflict, and evolutionary innovation. **Proceedings of the National Academy of Sciences**, v. 108, n. Supplement\_2, p. 10863–10870, 2011.
- WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973–982, 2007.
- WILDER, J.; HOLLOCHER, H. Mobile elements and the genesis of microsatellites in Dipterans. **Molecular Biology and Evolution**, v. 18, n. 3, p. 384–392, 2001.
- YUAN, Y. W.; WESSLER, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 19, p. 7884–7889, 2011.
- 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, n. 5, p. 584–596, 2013.

ZHOU, L. et al. Transposition of hAT elements links transposable elements and V(D)J recombination. **Nature**, v. 432, n. 7020, p. 995–1001, 2004.

ZIEMNICZAK, K. et al. In situ Localization of ( GATA ) n and ( TTAGGG ) n Repeated DNAs and W Sex Chromosome Differentiation in Parodontidae ( Actinopterygii : Characiformes ). **Cytogenetic and Genome Research**, v. 144, n. 4, p. 325–332, 2015.