

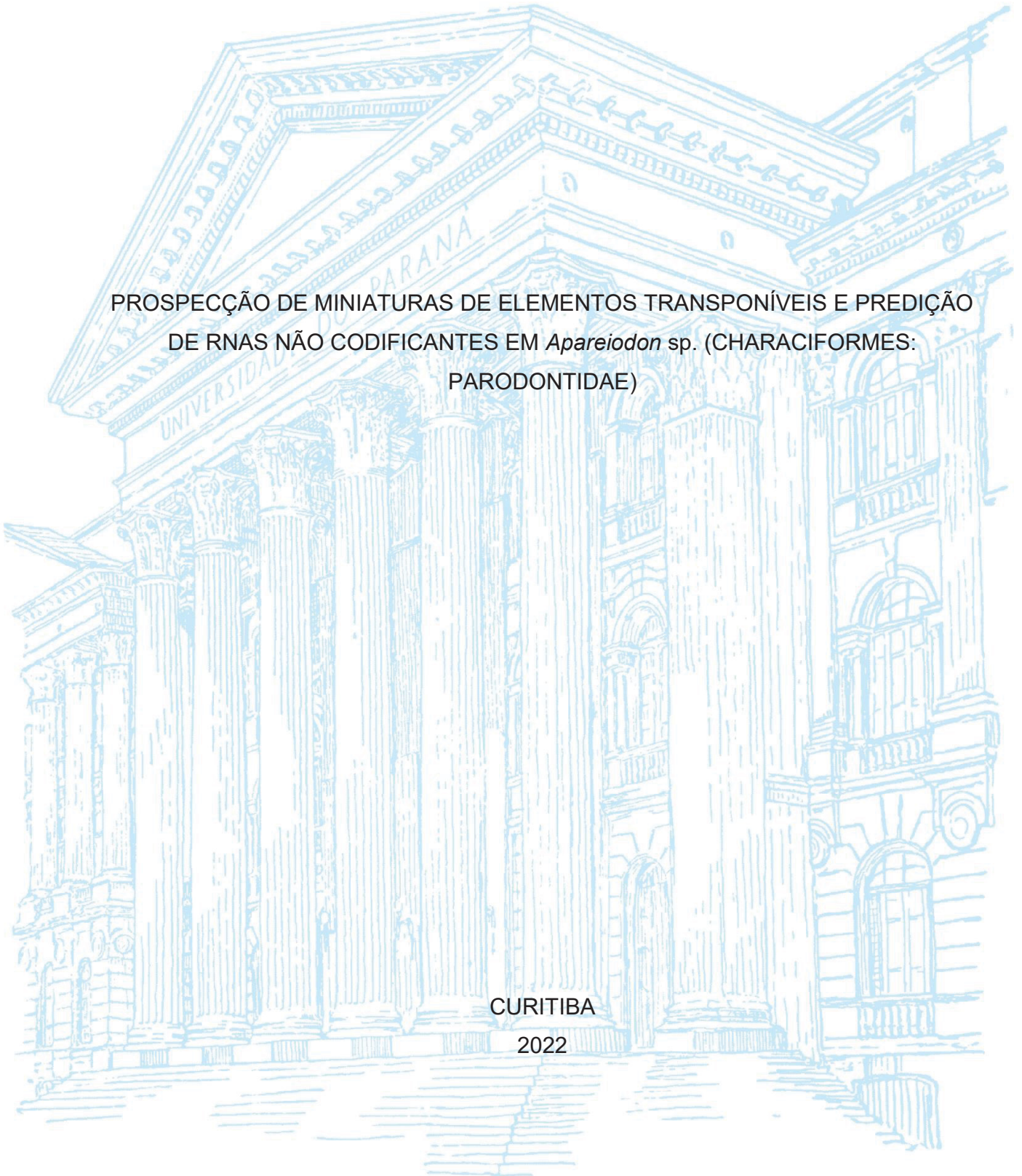
UNIVERSIDADE FEDERAL DO PARANÁ

MATHEUS AZAMBUJA DOS SANTOS

PROSPECÇÃO DE MINIATURAS DE ELEMENTOS TRANSPONÍVEIS E PREDIÇÃO
DE RNAs NÃO CODIFICANTES EM *Apareiodon* sp. (CHARACIFORMES:
PARODONTIDAE)

CURITIBA

2022



MATHEUS AZAMBUJA DOS SANTOS

PROSPECÇÃO DE MINIATURAS DE ELEMENTOS TRANSPONÍVEIS E PREDIÇÃO
DE RNAs NÃO CODIFICANTES EM *Apareiodon* sp. (CHARACIFORMES:
PARODONTIDAE)

Tese apresentada ao Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas da Universidade Federal do Paraná (UFPR) como parte dos requisitos para obtenção do título de Doutor em Ciências Biológicas, área de concentração Genética.

Orientador: Prof. Dr. Marcelo Ricardo Vicari
Coorientadora: Dr^a. Michelle Orane Schemberger

CURITIBA

2022

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

Santos, Matheus Azambuja dos.

Prospecção de miniaturas de elementos transponíveis e predição de RNAs não codificantes em *Apareiodon* sp. (Characiformes: Parodontidae). / Matheus Azambuja dos Santos. – Curitiba, 2022.

1 recurso on-line : PDF.

Orientador: Marcelo Ricardo Vicari.

Coorientadora: Michelle Orane Schemberger.

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

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



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

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **MATHEUS AZAMBUJA DOS SANTOS** intitulada: **Prospecção de miniaturas de elementos transponíveis e predição de RNAs não codificantes em *Apareiodon sp.* (Characiformes: Parodontidae)**, sob orientação do Prof. Dr. MARCELO RICARDO VICARI, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor 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, 12 de Setembro de 2022.

Assinatura Eletrônica

20/09/2022 16:45:22.0

MARCELO RICARDO VICARI

Presidente da Banca Examinadora

Assinatura Eletrônica

27/09/2022 19:50:34.0

PATRÍCIA PASQUALI PARISE MALTEMPI

Avaliador Externo (UNIVERSIDADE ESTADUAL PAULISTA - CAMPUS DE RIO CLARO)

Assinatura Eletrônica

21/09/2022 11:30:24.0

MARIA CLAUDIA GROSS

Avaliador Externo (UNIVERSIDADE FEDERAL DA INTEGRAÇÃO LATINO-AMERICANA - FOZ DO IGUAÇU)

Assinatura Eletrônica

20/09/2022 16:35:58.0

ADRIANA LUDWIG

Avaliador Externo (UNIVERSITY OF LIVERPOOL)

Assinatura Eletrônica

20/09/2022 17:09:50.0

MICHELLE ORANE SCHEMBERGER

Coorientador(a) (null)

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: 223760

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 223760

Dedico este trabalho aos meus pais, por sempre acreditaram em mim.

AGRADECIMENTOS

Agradeço primeiramente à Deus, por ter me dado forças e fé para a realização deste trabalho, mesmo nos momentos mais difíceis.

Aos meus pais, Airton e Adriana, que nunca mediram esforços para que seus filhos estudassem e corresse atrás dos seus objetivos. Por entenderem as longas ausências e os telefonemas cansados. Ao meu irmão, Bruno, que sempre emprestou seu apartamento em Curitiba quando eu precisei.

Ao meu orientador professor Marcelo Vicari, primeiramente por ter aceitado me orientar, e por sempre acreditar na minha capacidade. Obrigado pela oportunidade de ter trabalhado com você, pelos ensinamentos, conselhos, pela confiança em mim depositada, pela amizade, pelas piadas e brincadeiras, e pelos churrascos.

À Michelle Orane Schemberger, minha coorientadora, que sempre esteve disposta a ajudar e quem me auxiliou inicialmente com as análises de bioinformática.

À professora Viviane Nogaroto Vicari, pelo auxílio no laboratório, dicas e pela amizade.

Às professoras Kaline Ziemniczak e Adriana Ludwig pelas correções e sugestões durante o desenvolvimento deste trabalho.

Ao professor Cesar Martins e demais membros do projeto temático *Sex and B chromosomes enigmas: model system for study of chromosome and genome Evolution* pelas trocas de conhecimento e pelos momentos de descontração durante as reuniões e eventos do projeto.

Ao professor Orlando Moreira-Filho pela coleta dos peixes utilizados neste e em outros trabalhos, e pelas conversas sempre bem humoradas e descontraídas.

Ao professor David de Souza Jaccoud Filho e membros do laboratório do laboratório de Fitopatologia Aplicada da UEPG pela disponibilidade dos equipamentos. Aos professores José Rosa Gomes e Maria Albertina de Miranda Soares do laboratório de Biologia Celular, Molecular e do Desenvolvimento pela disponibilização do fotomicroscópio de epifluorescência para captura das imagens de FISH.

Aos colegas do CBSF Lab Viviane, Lucas, Larissa, Geize, Caroline, Sebastião, Thaís, Fernanda, Stephanie e Rafael pela amizade e apoio na rotina do laboratório.

Ao Programa de Pós-graduação em Genética, à Universidade Federal do Paraná e à Universidade Estadual de Ponta Grossa por toda a infraestrutura disponibilizada.

As agências de fomento CNPq, SETI, Fundação Araucária e FAPESP pelo apoio financeiro para realização da pesquisa, e à CAPES pela bolsa de estudo.

Por fim, agradeço a todos que de alguma forma contribuíram para a realização deste trabalho.

“Tubarão nada com Tubarão, sardinha nada com sardinha”

Autor desconhecido

*“Look, if you had one shot or one opportunity
To seize everything you ever wanted in one moment
Would you capture it, or just let it slip?*

...

*You only get one shot, do not miss your chance to blow
This opportunity comes once in a lifetime”*

Lose yourself – Eminem, Luis Resto e Jeff Bass

RESUMO

Genomas de eucariotos são compostos por diferentes classes de DNA repetitivo, como os elementos transponíveis (TEs) e as famílias multigênicas. Essa parcela do genoma é responsável por grande parte das variações cromossômicas e novidades evolutivas encontradas no DNA. Parodontidae é uma pequena família de peixes neotropicais, considerada um bom modelo para estudos evolutivos, uma vez que diferentes sistemas de cromossomos sexuais são descritos nela. A obtenção dos dados de sequenciamento de nova geração para a espécie *Apareiodon* sp. (Rio Verde), permitiram novas investigações para o entendimento da distribuição e atuação dos elementos repetitivos na espécie e para a família Parodontidae. Neste estudo, miniaturas de elementos transponíveis (MITEs), um grupo especial de pequenas sequências de TEs não autônomos, foram prospectadas e caracterizadas no genoma de *Apareiodon* sp. Em adição, pequenos RNAs nucleares (snRNAs) foram preditos e caracterizados em *Apareiodon* sp. Cinco U snRNAs identificados foram ainda localizados *in situ* nos cromossomos de espécimes dos gêneros *Apareiodon* e *Parodon*. A prospecção das MITEs identificou quase 10.000 sequências, e a caracterização destas demonstrou que cerca de 65% são similares à TEs das superfamílias *Tc-1/Mariner* e *hAT*, duas das superfamílias envolvidas em diferentes fases de invasão no genoma de *Apareiodon* sp. Muitas das MITEs foram localizadas próximas ou dentro de genes (introns e éxons), as qualificando para processos de co-opção molecular na espécie, o que foi confirmado pela identificação de famílias de pre-miRNAs e de miRNAs maduros por similaridade nas MITEs. A predição e caracterização dos U snRNAs demonstrou a presença de cópias putativamente funcionais para as nove famílias multigênicas envolvidas no *major* e *minor* spliceossomo (U1, U2, U4, U5, U6, U11, U12, U4atac e U6atac) no genoma de *Apareiodon* sp., em adição a muitas cópias defectivas e pseudogênicas. A localização *in situ* dos snDNA U1, U2, U4, U5 e U6 nas espécies de Parodontidae demonstraram o envolvimento dos snDNAs U2 e U4, em eventos independentes, na diferenciação dos cromossomos sexuais entre as linhagens da família.

Palavras-chave: MITEs; snRNAs; co-opção molecular; citogenética comparativa; cromossomos sexuais.

ABSTRACT

Eukaryotic genomes are composed of different classes of repetitive DNA, such as transposable elements (TEs) and gene families. This portion of the genome is responsible for a large portion of the chromosomal variations and evolutionary novelties found in DNA. Parodontidae is a small Neotropical fish family, considered a good model for evolutionary studies, once different systems of sex chromosomes are described. The new generation sequencing data obtained for the species *Apareiodon* sp. (Rio Verde), allowed further investigations to understand the distribution and performance of repetitive elements in the species and for the Parodontidae. In this study, Miniature Inverted-repeat Transposable Elements (MITEs), a special group of short sequences of non-autonomous TEs, were prospected and characterized in the genome of *Apareiodon* sp. In addition, small nuclear RNAs (snRNAs) were predicted and characterized in *Apareiodon* sp. Five identified U snRNAs were also located in situ in the chromosomes of *Apareiodon* and *Parodon*. The prospection of MITEs identified almost 10,000 sequences. Their characterization showed that about 65% are similar to the *Tc-1/Mariner* and *hAT* TEs, two of the superfamilies involved in different phases of invasion in the *Apareiodon* sp. genome. Many MITEs were located close to or within genes (introns and exons), qualifying them for molecular cooption processes in the species, which was confirmed identifying of families of pre-miRNAs and mature miRNAs by similarity in MITEs sequences. The prediction and characterization of U snRNAs demonstrated the presence of putatively functional copies for the nine gene families involved in the major and minor spliceosome (U1, U2, U4, U5, U6, U11, U12, U4atac, and U6atac) in the *Apareiodon* sp. genome, in addition to many defective and pseudogene copies. The in situ localization of U1, U2, U4, U5, and U6 snDNA in Parodontidae species demonstrated the involvement of the snDNAs U2 and U4, in independent events, in the differentiation of sex chromosomes among the Parodontidae lineages.

Keywords: MITEs; snRNAs; molecular co-option; comparative cytogenetic; sex chromosomes.

LISTA DE ILUSTRAÇÕES

MATERIAL E MÉTODOS

Figura 1 – Mapa do Brasil com a localização dos pontos de coleta dos espécimes de Parodontidae analisados.36

CAPÍTULO 1

Figura 1 – Identificação do local das MITEs no genoma de *Apareiodon* sp.62

Figura 2 – Identificação da similaridade das famílias de ncRNAs para as MITEs de *Apareiodon* sp. Nove MITEs apresentaram similaridade para duas ou mais famílias ncRNA.63

CAPÍTULO 2

Figure 1 – Prediction of the secondary structures of U snRNAs analyzed in the *Apareiodon* sp. genome.82

Figure 2 – Prediction of the U4/U6 and U4atac/U6atac snRNAs complexed structures of the *Apareiodon* sp. In (a) U4/U6 snRNAs complexed structure; and in (b) U4atac/U6atac snRNAs complexed structure.83

Figure 3 – Identification of the consensus regulatory sequences upstream to snRNAs transcribed regions in *Apareiodon* sp. genome. In (a – i) PSE; and in (j – k) TATA-box consensus.84

Figure 4 – In situ localization of the snDNAs sites on the Parodontidae species karyotypes. In (a, b, c, d, e), karyotypes of the *Apareiodon affinis*, *Apareiodon piracicabae*, *Apareiodon* sp., *Parodon hilarii*, and *Parodon nasus* species, respectively, submitted to FISH using U1 snDNA probe. In the boxes for each karyotype, respective chromosome bearing U2 snDNA sites (f, g, h, i, j); U4 snDNA sites (k, l, m, n, o); U5 snDNA sites (p, q, r, s, t); and U6 snDNA sites (u, v, w, x, y). Bar: 10 µm.85

CAPÍTULO 3

Figure 1 – Karyotypes of *Apareiodon affinis* populations (**a, b, c**), *A. ibitiensis* populations (**d, e**), *A. vittatus* (**f**), *A. vladii* (**g**), *Parodon buckleyi* (**h**), *P. pongoensis* (**i**), and *Parodon* sp. (**j**) submitted to FISH using U2 and U4 snDNAs probes. Bar = 10 μ m. 101

Figure 2 – Origin and differentiation of the ZZ/ZW₁W₂ sex chromosomes system in *Apareiodon affinis* from the Upper Paraná River. 102

LISTA DE TABELAS

MATERIAL E MÉTODOS

Tabela 1 – Espécies da família Parodontidae estudadas, localização, número de indivíduos utilizados e código de depósito.34

CAPÍTULO 1

Tabela 1 – Classificação das sequências MITEs identificadas no genoma de *Apareiodon* sp.57

Tabela 2 – Famílias de ncRNAs preditas nas sequências MITEs de *Apareiodon* sp.58

Tabela 3 – miRNAs identificados em sequências de MITEs e sua localização no genoma de *Apareiodon* sp.59

CAPÍTULO 2

Supplementary table 1 – Primers set designed and used to confirm snRNAs in *Apareiodon* sp. genome.86

Supplementary table 2 – Parodontidae species, sex chromosome systems and samples sites.87

Supplementary table 3 – Characterization of snRNAs amplicons obtained from *Apareiodon* sp.88

CAPÍTULO 3

Table 1 – Species of Parodontidae analyzed and sample location.99

Table 2 – *Apareiodon* and *Parodon* cytogenetic data and snDNAs sites. m: metacentric; sm: submetacentric; st: subtelocentric; p: short arm; q: long arm; PS: present study. .100

LISTA DE ABREVIATURAS

2n – Número diploide

a – Cromossomo acrocêntrico

AP – *Aspartic proteinase*

B – Cromossomo supranumerário

BLAST – *Basic Local Alignment Search Tool*

circRNAs – RNA circular

CTAB - *Cetyltrimethylammonium*

DSE – Sequência do elemento distal (*Distal Sequence Element*)

dsRNA – *Double strand RNA*

DW – *Downstream*

eRNA – *Enhancer RNA*

EX – Exônica

FISH – Hibridação *in situ* Fluorescente

gag – *Capsid protein*

IN – Intrônica

Int – *Initiator*

INT – *Integrase*

KCl – Cloreto de potássio

LINE – *Long Interspersed Nuclear Element*

lncRNAs – Longo RNA não codificante (*long non-coding RNA*)

LTR – Repetições terminais longas (*Long Terminal Repeat*)

m – Cromossomo metacêntrico

MgCl₂ – Cloreto de magnésio

miRNA – microRNA

MITE – Miniatura de elemento transponível (*Miniature Inverted-repeat Transposable Element*)

mM – Milimolar

MNRJ – Museu Nacional do Rio de Janeiro

μM – Micromolar

ncRNA – RNA não codificante (*non-coding RNA*)
ng – Nanograma
Non-LTR – Ausência de repetições terminais longas
nt – Nucleotídeos
NUPELIA – Núcleo de Pesquisas em Limnologia, Ictiologia e Aquacultura
OCT – *Octamer*
ORF – Fase de leitura aberta (*Open reading frame*)
p – Braço cromossômico curto
pb – Pares de bases
PCR – Reação em cadeia da polimerase (*Polymerase Chain Reaction*)
piRNA – piwi-RNA
pré-mRNAs – Precusores de RNAs mensageiros
PS: *Present study*
PSE – Sequência do elemento proximal (*Proximal Sequence Element*)
q – Braço cromossômico longo
RH – *RNAse H*
RNP – Ribonucleoproteína
rRNA – RNA ribossômico
RT – *Reverse transcriptase*
SINE – *Short Interspersed Nuclear Element*
siRNA – Pequeno RNA de interferência (*small interference RNA*)
sm – Cromossomo submetacêntrico
sncRNA – Pequeno RNA não codificante (*small non-coding RNA*)
snDNA – Pequeno DNA nuclear (*small nuclear DNA*)
snoRNA – Pequeno RNA nucleolar (*small nucleolar RNA*)
snRNA – Pequeno RNA nuclear (*small nuclear RNA*)
st – Cromossomo subtelocêntrico
TEs – Elementos transponíveis (*transposable elements*)
TIR – Repetição terminal invertida (*Terminal Inverted Repeat*)
tRNA – RNA de transferência
TSD – Sítios alvos de duplicação (*Target site duplication*)

U – Unidade

UP – *Upstream*

UTR – Região não traduzida (*untranslated region*)

YR – Tirosina recombinase

SUMÁRIO

1 INTRODUÇÃO	18
1.1 DNA REPETITIVO	18
1.1.1 Famílias multigênicas	18
1.1.2 Elementos transponíveis.....	21
1.2 RNAs NÃO CODIFICANTES	26
1.3 FAMÍLIA PARODONTIDAE	28
2 OBJETIVOS	33
2.1 OBJETIVO GERAL	33
2.2 OBJETIVOS ESPECÍFICOS	33
3 MATERIAL E MÉTODOS	34
3.1 MATERIAL BIOLÓGICO	34
3.2 MÉTODOS	37
3.2.1 Sequenciamento e montagem do genoma de <i>Apareiodon</i> sp.	37
3.2.2 Prospecção e caracterização de <i>Miniature Inverted-repeats Transposable Elements</i> no genoma de <i>Apareiodon</i> sp.....	38
3.2.3 Identificação de ncRNAs nas MITEs.....	39
3.2.4 Predição e caracterização de snRNAs em <i>Apareiodon</i> sp.	40
3.2.5 Identificação e caracterização de regiões promotoras dos snRNAs	41
3.2.6 Desenho dos <i>primers</i> e confirmação dos snRNAs no genoma de <i>Apareiodon</i> sp.....	41
3.2.7 Hibridação <i>in situ</i> Fluorescente (FISH)	42
4 RESULTADOS	43
4.1 CAPÍTULO 1	44

Miniaturas de elementos transponíveis no genoma de <i>Apareiodon</i> sp. (Characiformes: Parodontidae).....	44
4.2 CAPÍTULO 2	64
Major and minor U small nuclear RNAs genes characterization in a neotropical fish genome: Chromosomal remodeling and repeats units dispersion in Parodontidae.....	64
4.3 CAPÍTULO 3	89
Comparative U2 and U4 snDNA chromosomal mapping in the Neotropical fish genera <i>Apareiodon</i> and <i>Parodon</i> (Characiformes: Parodontidae).....	89
5 CONSIDERAÇÕES FINAIS	103
6 REFERÊNCIAS.....	104

1 1 INTRODUÇÃO

2

3 1.1 DNA REPETITIVO

4

5 Os cromossomos eucarióticos são constituídos de diferentes classes de DNA, as
6 quais podem ser encontradas uma única vez no genoma (DNA cópia simples), ou
7 sequências que se repetem de dezenas a milhões de vezes (DNA repetitivo) (SUMNER,
8 2003). As sequências de DNAs repetitivos podem ainda ser classificadas como famílias
9 multigênicas (DNA ribossômico, histonas, snRNAs, globinas), sequências dispersas,
10 comumente identificadas como elementos transponíveis (TEs), e repetidas *in tandem*
11 (satélites, minissatélites e microssatélites) (SUMNER, 2003, LÓPEZ-FLORES;
12 GARRIDO-RAMOS, 2012).

13 Os DNAs repetitivos compreendem uma grande porção dos genomas
14 eucarióticos, tendo funções estruturais e organizacionais nesses, e são conhecidos por
15 estarem envolvidos em mecanismos que geram diferenciação e rearranjos
16 cromossômicos (BISCOTTI; OLMO; HESLOP-HARRISON, 2015).

17

18 1.1.1 Famílias multigênicas

19

20 Famílias multigênicas são descritas como um grupo de genes com função similar
21 e com sequências que se originaram de um gene ancestral comum (NEI; ROONEY,
22 2005). Os produtos destes genes, geralmente, são requeridos em grandes quantidades
23 nas células, e as múltiplas cópias tendem a ser originadas por processos de duplicação
24 gênica (SUMNER, 2003). No entanto, outros fenômenos como pseudogenização, perda
25 de genes, recombinação e a seleção natural agem em diferentes níveis para moldar a
26 evolução das famílias gênicas (EIRÍN-LOPEZ et al., 2012).

27 Famílias multigênicas apresentam diferentes cenários em relação ao
28 mapeamento cromossômico nas espécies, estas podem estar localizadas em
29 cromossomo homeólogos em espécies relacionadas, bem como eventos de dispersão
30 destas ao longo dos cromossomos do complemento são observados (GORNUNG, 2013;
31 ANJOS et al., 2015; VITALES et al., 2017; SOCHOROVÁ et al., 2018; DEGRANDI et al.,

1 2020). Enquanto as famílias dos rRNAs 45S e 5S são amplamente mapeadas nos
2 cromossomos de peixes (GORNUNG, 2013; SOCHOROVÁ et al., 2018), estudos com
3 genes histonas e dos pequenos RNAs nucleares (snRNAs – *small nuclear RNAs*) ainda
4 são escassos, principalmente em espécies neotropicais (PAIM et al., 2018).

5 Inúmeros genes transcritos pela RNA polimerase II de eucarióticos são
6 expressos como mRNAs precursores (pré-mRNAs) que são convertidos em mRNAs por
7 *splicing*, uma etapa essencial da expressão gênica em que sequências não codificantes
8 (íntrons) são removidas e sequências codificantes (éxons) são unidas (WILL;
9 LÜHRMANN, 2011). A complexa maquinaria responsável pelo processo de *splicing* é
10 denominada de spliceossomo e envolve cinco diferentes subunidades de
11 ribonucleoproteínas (RNP) associadas a muitos cofatores de proteínas (LERNER et al.,
12 1980; WILL; LÜHRMANN, 2011, SHI, 2017). Dois tipos de spliceossomos são
13 encontrados em muitos eucariotos: o spliceossomo dependente de U-2 (*major*
14 spliceossomo), que catalisam íntrons do tipo U-2, e um menos abundante dependente de
15 U-12 (*minor* spliceossomo) que removem os raros íntrons do tipo U-12 (PATEL; STEITZ,
16 2003; TURUNEN et al., 2013). O *major* spliceossomo é composto pelos *snRNAs U1, U2,*
17 *U4, U5 e U6* e múltiplas proteínas (PATEL; STEITZ, 2003). Já o *minor* spliceossomo, o
18 qual além de ser organizado por múltiplas proteínas e o *snRNA U5* (comum ao *major*
19 spliceossomo), utiliza os *snRNAs U11, U12, U4atac e U6atac*, os quais possuem
20 estrutura e função similares aos *snRNAs U1, U2, U4 e U6* (PATEL; STEITZ, 2003).
21 Íntrons do tipo U-12 representam, geralmente, menos de 0.5% do total de íntrons
22 observados nos genomas (TURUNEN et al., 2013), e diferem de íntrons do tipo U-2 pela
23 terminação consenso que é reconhecida pelo spliceossomo: enquanto íntrons do tipo U-
24 2 possuem terminações GT-AG (raramente AT-AC e GC-AG), o tipo U-12 possui
25 terminações AT-AC (raramente GT-AG) (PATEL; STEITZ, 2003; SHETH et al., 2006).

26 Membros das famílias gênicas dos snRNAs são caracterizados por um arranjo
27 único de elementos promotores, que incluem um *Distal Sequence Element* (DSE), que
28 atua como um *enhancer*, e um *Proximal Sequence Element* (PSE). Alguns genes podem
29 apresentar ainda um TATA-box localizado adjacente ao PSE (HERNANDEZ, 2001;
30 JAWDEKAR; HENRY, 2008). A presença combinada dos elementos PSE e TATA
31 direciona para o recrutamento do mecanismo específico da RNA polimerase III, enquanto

1 a ausência de TATA-box leva ao recrutamento do aparato específico de transcrição da
2 RNA polimerase II (HERNANDEZ, 2001; JAWDEKAR; HENRY, 2008). DSE e PSE
3 podem ser intercambiados entre as RNAs polimerases II e III sem efeitos na
4 especificidade das RNAs polimerases, a qual é determinada pela presença ou ausência
5 do TATA-box (HERNADEZ, 2001).

6 Estudos de predição e caracterização de snRNAs em peixes são escassos:
7 López et al. (2008) observaram a ocorrência das nove famílias de snRNAs no genoma
8 das espécies de peixes *Fugu rubripes*, *Gasterosteus aculeatus*, *Oryzias latipes*,
9 *Petromyzon marinus* e *Tetraodon nigroviridis*, enquanto que em *Danio rerio* o *snRNA U12*
10 não foi predito; Já Marz, Kirsten e Stadler (2008) quando analisaram as mesmas 6
11 espécies, identificaram todos os snRNAs funcionais, com exceção do *snRNA U4atac* em
12 *T. nigroviridis*.

13 Com relação à localização *in situ* destes genes, U snDNAs são considerados
14 bons marcadores para estudos evolutivos entre espécies relacionadas, para inferir a
15 homeologia dos cromossomos, e para rastrear a origem e evolução de certos
16 cromossomos, como os cromossomos supranumerários (Bs) e sexuais (UTSUNOMIA et
17 al., 2014). No entanto, de maneira geral os estudos focam na localização dos snDNAs
18 U1 e U2.

19 Para o snDNA U1, peixes geralmente apresentam um único locus no cariótipo
20 (CABRAL-DE-MELLO et al., 2012; MALIMPENSA et al., 2020; YANO et al., 2020;
21 SOARES et al., 2021). No entanto, sítios múltiplos ou sinais dispersos já foram descritos
22 (MANCHADO et al., 2006; GARCÍA-SOUTO et al., 2015; SILVA et al., 2015; DULZ et al.,
23 2020).

24 Para o snDNA U2, a condição de locus simples também é a mais descrita
25 (MERLO et al., 2010, 2012a, 2013; SUPIWONG et al., 2013; UTSUNOMIA et al., 2014;
26 GARCÍA-SOUTO et al., 2015; ARAYA-JAIME et al., 2017; SERRANO et al., 2017;
27 PISCOR; FERNANDES; PARISE-MALTEMPI, 2018; PONZIO; PISCOR; PARISE-
28 MALTEMPI, 2018; PUCCI et al., 2018; MALIMPENSA et al., 2018, 2020; SEMBER et al.,
29 2018; DULZ et al., 2020; OLIVEIRA et al., 2020; SOARES et al., 2021; HAERTER et al.,
30 2022). Em alguns casos, esse snDNA pode estar presente em múltiplos sítios
31 (MANCHADO et al., 2006; UTSUNOMIA et al., 2014; SILVA et al., 2015; PISCOR;

1 CENTOFANTE; PARISE-MALTEMPI, 2016; YANO et al., 2017) ou disperso ao longo dos
2 cromossomos (ÚBEDA-MANZANARO et al., 2010; MERLO et al., 2012b; RÁB et al.,
3 2016; HATANAKA et al., 2018).

4 Os snDNAs U4 e U6 foram mapeados apenas na espécie *Hollandichthys*
5 *multifasciatus* (SOARES et al., 2021), e sítios únicos independentes para estes snDNAs
6 foram encontrados. Para o snDNA U5, não existem relatos na literatura de estudos de
7 localização *in situ* deste elemento em peixes.

8 Sinais de snDNAs foram descritos nos cromossomos sexuais de *Gymnotus*
9 *pantanal* (UTSUNOMIA et al., 2014) e em *Triportheus albus* (YANO et al., 2017), e
10 parecem ser eventos espécie-específicos destes cromossomos nas espécies.

11 De acordo com Marz, kirsten e Stadler (2008), snRNAs se comportam como
12 elementos transponíveis, uma vez que há pouca conservação quando os genes
13 flanqueadores destas sequências são analisados em diferentes genomas. Associação de
14 snRNAs com genes de rRNAs (MANCHADO et al., 2006; ÚBEDA-MANZANARO et al.,
15 2010; SCACCHETTI et al., 2015; SILVA et al., 2015; YANO et al., 2017, 2020;
16 MALIMPENSA et al., 2018, 2020; PISCOR; FERNANDES; PARISE-MALTEMPI, 2018),
17 em sintenia a outros snRNAs (MANCHADO et al., 2006; MARZ; KIRSTEN; STADLER,
18 2008; MERLO et al., 2010, 2012a, 2013; GARCÍA-SOUTO et al., 2015) ou com genes
19 histônicos (MALIMPENSA et al. 2018) já foram descritos.

20

21 1.1.2 Elementos transponíveis

22

23 Elementos repetitivos dispersos, também conhecidos como elementos genéticos
24 móveis, ou elementos transponíveis, são segmentos de DNA que possuem a capacidade
25 de se replicarem e dispersarem nos genomas hospedeiros (WICKER et al., 2007;
26 WELLS; FESCHOTTE, 2020). São conhecidos por ocupar uma fração considerável
27 dentro dos genomas eucariotos, e devido as diversas estratégias de invasão de novas
28 regiões representam a maior fonte de variação e novidades genéticas (WELLS;
29 FESCHOTTE, 2020). Os TEs foram identificados na década de 40, por Bárbara
30 McClintock, denominando-os como “elementos controladores”, sugerindo que tais

1 elementos desempenhavam um papel relevante na regulação gênica (MCCLINTOCK,
2 1950).

3 Os TEs são agrupados em duas maiores classes, de acordo com o seu
4 mecanismo de transposição: Classe I ou retrotransposons, os quais se transpõem
5 através de um intermediário de RNA pelo mecanismo de transcrição reversa, e Classe II
6 ou transposons de DNA, os quais possuem capacidade de excisão do seu local de origem
7 e inserção em outro lugar do genoma (FINNEGAN, 1989). Ambas as classes podem
8 ainda ser divididas em subclasses e ordens de acordo com a estrutura molecular dos
9 elementos ou pelo número de fitas excisadas durante a transposição (WICKER et al.,
10 2007). Posteriormente são agrupados em superfamílias de acordo com a organização
11 das suas fases aberta de leitura (ORFs – *Open reading frame*), presença ou ausência de
12 domínios e assinatura proteica, e finalmente em famílias e subfamílias, principalmente no
13 que diz respeito à similaridade entre as sequências (WICKER et al., 2007).

14 Retrotransposons podem ser divididos em três maiores subclasses, de acordo
15 com o mecanismo de replicação e integração: (I) elementos com repetições terminais
16 longas (LTR – *Long Terminal Repeat*); (II) elementos sem repetições terminais longas
17 (non-LTR); e (III) elementos mobilizados por Tirocinas recombinases (YR) (WELLS;
18 FESCHOTTE, 2020). Elementos LTR geralmente codificam uma proteína gag (*capsid*
19 *protein*) e as enzimas AP (*aspartic proteinase*), RT (*reverse transcriptase*), RH (*RNAse*
20 *H*), e INT (*integrase*), movendo-se no genoma pelo mecanismo de transcrição em RNAm,
21 conversão em cDNA pela enzima RT, seguido da inserção em uma nova região do
22 genoma pela *integrase* (WICKER et al., 2007; WELLS; FESCHOTTE, 2020). Nos
23 Elementos non-LTR, durante o processo de transposição, o RNA intermediário é
24 inicialmente ancorado ao DNA alvo, este é então transcrito reversamente, e a nova fita
25 de cDNA é integrada ao DNA (WICKER et al., 2007; WELLS; FESCHOTTE, 2020). Ainda,
26 para os elementos YR, é proposto que durante a transposição destes, após a transcrição
27 reversa do RNAm, este é circularizado para que ocorra a síntese da segunda fita do
28 cDNA, e então integrado à nova região do DNA por uma tirosina recombinase
29 (POULTER; BUTLER, 2015). Devido à característica de transposição dos
30 retrotransposons, estes são popularmente conhecidos como elementos “copia e cola”
31 (WICKER et al., 2007).

1 TEs da classe II são caracterizados por se moverem diretamente de DNA para
2 DNA, sendo divididos em duas subclasses: os elementos conhecidos como “corta e cola”,
3 que podem ser mobilizados por transposases (ordem TIR – *Terminal Inverted Repeat*),
4 ou por tirosinas recombinases (*Cryptons*); na subclasse II encontram-se os elementos
5 que são transpostos por um mecanismo de “copia e cola”, que compreendem os
6 elementos círculo-rolante (conhecidos como *Helitrons*), e os transposons “*self-*
7 *synthesizing*”, conhecidos como *Mavericks* e *Polintons* (WICKER et al., 2007; WELLS;
8 FESCHOTTE, 2020). Para a transposição dos elementos da ordem TIR ocorre a excisão
9 dos dois filamentos da fita de DNA. Elementos da ordem TIR se caracterizam por
10 apresentarem um gene para a enzima transposase (KOJIMA, 2019). O mecanismo
11 preciso de ação das transposases varia entre as superfamílias, no entanto, inicia pelo
12 corte próximo a região terminal de cada TIR, removendo o transposon de DNA do seu
13 local de origem, e inserindo a dupla fita de DNA em uma nova região a partir de um corte
14 coesivo nesta, originando sítios alvos de duplicação (TSD – *Target site duplication*)
15 (HICKMAN; DYDA, 2016). Apesar deste processo ser não replicativo, estes elementos
16 podem aumentar seu número de cópias no genoma, realizando a transposição do
17 elemento durante a replicação do DNA, ou por reparo, por recombinação homóloga,
18 durante a quebra da dupla de DNA (WICKER et al., 2007; WEELS; FESCHOTTE, 2020).

19 Já durante a transposição dos elementos da subclasse II, pode haver a replicação
20 de uma ou duas fitas de DNA para sua posterior transferência (WICKER et al., 2007).
21 Para esses TEs, proteínas específicas realizam o corte de apenas uma fita do DNA,
22 seguido de síntese da fita complementar e inserção do elemento em uma nova região
23 genômica (WICKER et al., 2007; WELLS; FESCHOTTE, 2020).

24 Elementos transponíveis das classes I e II podem ainda ser classificados como
25 elementos autônomos, quando apresentam todos os domínios íntegros para sua
26 transposição, ou não autônomos, quando dependem das enzimas de outros elementos
27 para se moverem no genoma (WICKER et al., 2007). A grande maioria das famílias de
28 DNA apresentam cópias de elementos não autônomos, as quais são geradas pela
29 degeneração ou deleção das ORFs, mantendo apenas as regiões terminais do TE
30 original, as quais ainda podem ser reconhecidas pelas enzimas de elementos

1 relacionados e continuarem a se dispersar pelo genoma hospedeiro (KOJIMA, 2019;
2 WELLS; FESCHOTTE, 2020).

3 O “ciclo de vida” dos TEs consiste de uma série de etapas a partir do seu
4 “surgimento” no genoma até sua “morte”, as quais incluem o processo de invasão no
5 genoma hospedeiro, uma intensa fase de replicação com aumento do número de cópias
6 destes, uma fase de inativação, e sua eliminação (KIDWELL e LISCH, 1997;
7 FERNÁNDEZ-MEDINA et al., 2012). Após sua inativação, os elementos
8 progressivamente acumulam mutações, e inserções e deleções (*indels*) em suas
9 sequências, até perderem completamente suas identidades (FERNÁNDEZ-MEDINA et
10 al., 2012).

11 Miniaturas de elementos transponíveis (MITes – *Miniature Inverted-repeat*
12 *Transposable Elements*) são transposons de DNA não autônomos, caracterizados pelo
13 alto número de cópias no genoma, e com tamanho relativamente pequeno (50 – 800 pb)
14 (WICKER et al., 2007; FATTASH et al., 2013; CRESCENTE et al., 2018). MITes possuem
15 TIRs e podem apresentar ainda duas repetições flanqueadoras curtas, as TSD, geradas
16 durante o processo de transposição (WICKER et al., 2007; CRESCENTE et al., 2018).
17 MITes frequentemente são encontradas próximas ou dentro de genes, onde podem
18 afetar sua expressão (PIRIYAPONGSA; JORDAN, 2008; KUANG et al., 2009; KLAI et al.,
19 2020; 2022; MACKO-PODGÓRNI; MACHAJ; GRZEBELUS, 2021).

20 MITes foram descobertas e descritas pela primeira vez no genoma de milho (*Zea*
21 *mays*) (BUREAU; WESSLER, 1992), e posteriormente foram identificadas em diferentes
22 grupos de organismos (FASTTASH et al., 2013; CHEN et al., 2014b; LIU et al., 2019;
23 ALBUQUERQUE; EBERT; HAAG, 2020; MIKKELSEN; WEIR, 2020; GRACE; CARR,
24 2020). No entanto, estudos de prospecção e identificação de MITes em genomas de
25 peixes ainda são escassos (IZSVÁK et al., 1999; SCHEMBERGER et al., 2016). Izsvák
26 et al. (1999) descreveram a primeira família de MITes isolada de peixes, a família *Angel*,
27 cópias descritas como pertencentes à mesma família foram encontradas no genoma de
28 *Danio rerio*, *D. albolineatus*, *D. aequipinnatus* e *Oryzias latipes*, e foram localizadas em
29 diferentes regiões gênicas. Já Schemberger et al. (2016) descreveram a presença de
30 cópias MITes com similaridade à *Tc1-Mariner* nas espécies *Parodon hiliarii* e *P.*
31 *pongoensis* (Parodontidae).

1 A origem e dispersão das MITEs nos genomas é descrita por ocorrer em duas
2 etapas a partir de Transposons de DNA: inicialmente, a transposição de elementos
3 autônomos dá origem à várias cópias derivadas não autônomas, com deleções nas
4 regiões internas. Em um segundo momento, é proposto que essas cópias derivadas
5 podem (por razões desconhecidas) amplificar seu número de cópias (FESCHOTTE;
6 JIANG; WESSLER, 2002). De acordo com este modelo, MITEs podem ser derivadas de
7 elementos autônomos ou elementos não autônomos por meio de mecanismos como
8 reparo abortivo de lacunas (FATTASH et al., 2013). Quando seus elementos autônomos
9 originais são perdidos ou inativados em um genoma, algumas MITEs podem ainda ser
10 mobilizadas de forma cruzada por elementos autônomos relacionados, devido à presença
11 de *motifs* (TIRs) conservados que permitem a transposição (FATTASH et al., 2013).

12 O número de MITEs e de famílias é uma característica intrínseca de cada espécie
13 (LIU et al., 2019; MACKO-PODGÓRNI; MACHAJ; GRZEBELUS, 2021; KLAI et al., 2022),
14 por exemplo em plantas, as MITEs podem representar grande fração dos genomas,
15 podendo chegar a mais de 200.000 cópias em algumas espécies (CHEN et al., 2014b),
16 no entanto, em alguns grupos animais um número menor de cópias são encontradas,
17 como nas borboletas *Helicoverpa armigera* e *Helicoverpa zea*, com 3.570 e 7.415
18 sequências, respectivamente (KLAI et al., 2020, 2022). Em algumas espécies ainda,
19 poucas sequências de MITEs podem ser encontradas, como observado na ave
20 rendalinho-do-xingu (*Willisornis vidua nigrigula*) com apenas 38 elementos (MIKKELSEN;
21 WEIR, 2020), ou estas podem estar ausentes do genoma, como descrito para a alga
22 vermelha *Cyanidioschyzon merolae* e no protozoário *Plasmodium falciparum*
23 (CRESCENTE et al., 2018).

24 O conceito de domesticação, co-opção ou exaptação molecular, foi descrito para
25 caracterizar o processo pelo qual uma sequência de TE é cooptada para realizar uma
26 nova função e trazer benefícios para o genoma (MILLER; MCDONALD; PINSKER, 1997).
27 Quando inseridas na região *upstream*, MITEs podem gerar novos sítios de ligação de
28 fatores de transcrição (TESTORI et al., 2012; HÉNAFF et al., 2014; MORATA et al.,
29 2018). Por outro lado, a presença de MITEs próximas à genes podem suprimir a
30 expressão destes, através da metilação do DNA (ARIEL; MANAVELLA, 2021). Devido à
31 presença das TIRs, uma vez transcritas, MITEs podem se dobrar em uma estrutura de

1 grampo (*stem*) similares as sequências de pré-miRNA processados pela enzima Dicer
2 para produzir miRNAs maduros (PIRIYAPONGSA; JORDAN, 2008). Piriyaopngsa e
3 Jordan (2008), descreveram a presença de 10 e 38 miRNAs derivados de sequências
4 MITEs nos genomas de *Arabidopsis thaliana* e *Oryza sativa*, respectivamente, já
5 Crescente et al. encontraram 38 miRNAs derivados de MITEs no genoma de trigo
6 (*Triticum aestivum*).

7

8 1.2 RNAs NÃO CODIFICANTES

9

10 RNAs não codificantes (ncRNA – *non-coding RNA*) representam a maior porção
11 dos transcritos nas células eucarióticas (AMARAL et al., 2008; LIU et al., 2013;
12 DEVESON et al., 2017). De maneira geral, ncRNAs participam de múltiplos processos
13 biológicos, da regulação fisiológica dos organismos e do processo de desenvolvimento,
14 ou mesmo de doenças (ZHANG et al., 2019).

15 A transcrição do DNA eucarioto de diferentes regiões genômicas e o
16 processamento do RNA produzem vários tipos de ncRNAs, os quais são classificados de
17 acordo com sua função regulatória (ZHANG et al., 2019). ncRNAs podem ser divididos
18 em duas categorias: *housekeeping* ncRNAs, os quais são abundantes nas células, e que
19 inicialmente regulam funções celulares genéricas, e os *regulatory* ncRNAs que atuam
20 principalmente como reguladores da expressão gênica em níveis epigenéticos,
21 transcricionais e pós-transcricionais (PONJAVIC et al., 2007; CECH; STEITZ, 2014;
22 PESCHANSKY; WAHLESTEDT, 2014; ZHANG et al., 2019). Entre os *housekeeping*
23 ncRNAs encontram-se os RNAs ribossômicos (rRNA), os RNAs de transferência (tRNA),
24 os pequenos RNAs nucleares (snRNA) e os pequenos RNAs nucleolares (snoRNAs). Já
25 nos *regulatory* ncRNAs estão agrupados os microRNAs (miRNA), os pequenos RNAs de
26 interferência (siRNA), os piwi-RNAs (piRNA), enhancer RNA (eRNA), os longos RNAs
27 não codificantes (lncRNAs) e os RNAs circulares (circRNAs) (ZHANG et al., 2019).

28 Housekeeping ncRNAs foram intensamente estudados nas últimas décadas,
29 incluindo rRNAs, tRNAs, snRNAs e snoRNAs. Esses ncRNAs são usualmente pequenos,
30 com tamanho que varia de 50 à 500 nucleotídeos, expressos constitutivamente em todos
31 os tipos celulares e necessários para a viabilidade celular (CHEN et al., 2018). Além da

1 função essencial dos rRNAs e tRNAs na síntese proteica, dos snRNAs no splicing e dos
2 snoRNAs na modificação de RNAs, housekeeping ncRNAs podem atuar como RNAs
3 regulatórios quando clivados (ZHANG et al., 2019).

4 Baseado no tamanho, os *regulatory* ncRNAs podem ser classificados em
5 pequenos RNAs não codificantes (sncRNAs – *small non-coding* RNAs), que possuem até
6 200 nucleotídeos, e lncRNAs, os quais possuem mais de 200 nucleotídeos (ZHANG et
7 al., 2019). Os principais representantes dos sncRNAs são os miRNAs, siRNAs e piRNAs,
8 no entanto, alguns ncRNAs com tamanhos variáveis podem pertencer às duas
9 classificações ao mesmo tempo, como os eRNAs e cicRNAs (ZHANG et al., 2019).

10 miRNAs é a classe mais abundante de sncRNAs, estes elementos são originados
11 a partir da formação de estruturas em *hairpin-loop* de sequências transcritas, os pre-
12 miRNAs, que após serem processados pela maquinaria do RNAi darão origem a
13 sequências de miRNAs maduras com 19 a 25 nucleotídeos (LU; ROTHENBERG, 2018).
14 miRNAs são reguladores endógenos que atuam no citoplasma e no núcleo celular
15 através de diferentes mecanismos, mediando o silenciamento gênico no nível pós-
16 transcricional (LU; ROTHENBERG, 2018; ZHANG et al., 2019).

17 siRNAs são uma classe de moléculas de RNAs dupla-fita que possuem
18 importante função em vias de RNAi, no entanto, diferente de miRNAs, que atuam na
19 degradação do RNAm ou repressão da tradução deste, siRNAs atuam apenas na
20 degradação dos RNAm (ZHANG et al., 2019; HU et al., 2020).

21 Identificados após a descoberta das proteínas PIWI, piRNAs são uma classe de
22 pequenos ncRNAs exclusiva de animais (ZHANG et al., 2019). piRNAs maduros são
23 derivados de sequências precursoras transcritas a partir de clusters de piRNAs que são
24 clivadas por proteínas PIWI (ARAVIN; HANNON; BRENNKE, 2007). piRNAs se
25 associam com as proteínas PIWI, formando um complexo de silenciamento piRNA-
26 induced, realizando o controle da expressão de seus alvos nos níveis transcricionais e
27 pós-transcricionais (SIOMI et al., 2011).

28 lncRNAs são definidos como transcritos que possuem mais de 200 nucleotídeos
29 e que não possuem informação para serem traduzidos (ZHANG et al., 2019). De acordo
30 com seus efeitos regulatórios estes podem ser classificados como cis-lncRNAs, que
31 regulam a expressão de genes próximos a sua localização no DNA, ou trans-lncRNAs

1 que regulam genes distantes (KORNIENKO et al., 2013). Alguns lncRNAs podem ainda
2 ser processados e dar origem à pequenos RNAs regulatórios, como miRNAs, piRNAs e
3 snoRNAs (RÖTHER; MEISTER, 2011).

4 eRNAs é um grupo emergente de ncRNAs, similares aos lncRNAs, no entanto
5 correspondem à transcritos das regiões regulatórias *enhancer* dos genes (ZHANG et al.,
6 2019). CircRNAs são uma classe única de ncRNAs que formam estruturas circulares
7 fechadas, estes apresentam tamanho variável (100 a 10.000 nucleotídeos), e são
8 originados a partir do splicing de éxons, introns, regiões intergênicas, e regiões não
9 traduzidas 5' e 3' (UTRs – *untranslated regions*) (JECK et al., 2013; YE et al., 2015; NOTO;
10 SCHMIDT; MATERA, 2017; ZHANG et al., 2019).

11

12 1.3 FAMÍLIA PARODONTIDAE

13

14 A família Parodontidae (Eigenmann, 1910), pertencente a ordem Characiformes,
15 é constituída por um pequeno grupo de peixes Neotropicais reunidos em 3 gêneros:
16 *Parodon* Valenciennes, 1849, *Saccodon* Kner, 1863 e *Apareiodon* Eigenmann, 1916
17 (PAVANELLI; BRITSKI, 2003). Atualmente são reconhecidas 32 espécies na família
18 (*Apareiodon* – 15 espécies; *Parodon* – 14 espécies; *Saccodon* – 3 espécies) (FRICKE;
19 ESCHMEYER; FONG, 2022). São peixes com formato fusiformes, com um colorido
20 cinza-claro e nadadeiras peitorais e pélvicas bem desenvolvidas, e comumente
21 denominados de “canivetes”, “peixe-charuto” ou “virolitos”, com pouco interesse
22 comercial, habitando ambientes lóticos de riachos de cabeceira de fundos rochosos
23 (NAKATANI et al., 2001; PAVANELLI, 2003). Espécimes de Parodontidae encontram-se
24 distribuídos por todo o continente Sul Americano e parte do Panamá, com exceção de
25 alguns rios da Bacia Atlântica, da Patagônia e no canal central do rio Amazonas
26 (PAVANELLI, 2003).

27

28 As espécies pertencentes à Parodontidae podem ser facilmente diagnosticadas
29 dos demais Characiformes por apresentarem boca inferior com mandíbula espatulada e
30 desprovida de dentes em sua porção anterior, dentes pré-maxilares pedunculados com
31 borda distal larga, multicuspidada e distribuídos em uma única série, e pela ausência de
lábio superior e fontanela (PAVANELLI, 2003; INGENITO, 2008). A caracterização

1 morfológica dos gêneros se dá pela variação de poucos caracteres: número de raios
2 indivisos nas nadadeiras peitorais (um raio em *Parodon* e *Apareiodon* e dois raios em
3 *Saccodon*), e ausência de dentes na região lateral da mandíbula em *Apareiodon* e
4 *Saccodon* (PAVANELLI; BRITSKI, 2003).

5 Do ponto de vista citogenético, as espécies dos três gêneros apresentam um
6 número diploide ($2n$) conservado de 54 cromossomos, compostos por cromossomos bi-
7 armados, com prevalência da ocorrência de cromossomos metacêntricos e
8 submetacêntricos, enquanto que cromossomos subtelocêntricos estão ausentes ou
9 variam de um a três pares entre as espécies (BELLAFRONTTE et al., 2011; 2012; TRALDI
10 et al., 2016; 2020; NIRCHIO et al., 2021). No entanto, populações de *Apareiodon affinis*
11 dos rios Cuiabá, Paraguai e Uruguai, podem apresentar de 2 à 8 pares de cromossomos
12 acrocêntricos (NASCIMENTO et al., 2018). Variação no $2n$ foram encontradas em
13 indivíduos de *Apareiodon piracicabae* do rio Passa Cinco (Ipeúna – São Paulo), e
14 *Parodon nasus* do rio Três Bocas (Londrina – Paraná) devido a presença de
15 cromossomos supranumerários (Bs) (FALCÃO; MOREIRA-FILHO; BERTOLLO, 1984;
16 PAULA et al., 2017). Ainda, as populações de *A. affinis* do Alto Paraná apresentam $2n =$
17 55 em fêmeas, devido ao sistema múltiplo de cromossomos sexuais (MOREIRA-FILHO;
18 BERTOLLO; GALETTI Jr, 1980).

19 Embora as espécies de Parodontidae apresentem um $2n$ conservado, variações
20 relacionadas à presença de cromossomos sexuais heteromórficos, localização
21 cromossômica dos marcadores rDNAs 18S e 5S, DNA satélite pPh2004, fração repetitiva
22 WAp, elementos transponíveis, histonas e pequenos RNAs nucleares (snRNA – *small*
23 *nuclear* RNA) são evidenciadas (BELLAFRONTTE et al., 2011, 2012; SCHEMBERGER
24 et al., 2011, 2014, 2016, 2019; ZIEMNICZAK et al., 2014; TRALDI et al., 2016, 2019,
25 2020; NASCIMENTO et al., 2018; SANTOS et al., 2019; AZAMBUJA et al., 2022a, b).

26 Com relação ao rDNA 18S, os sítios detectados pela técnica de impregnação por
27 nitrato de prata (Ag-NORs) e sondas de rDNA 18S, demonstraram a localização deste na
28 região terminal do braço longo de um par de cromossomos subtelocêntricos nas espécies
29 *A. affinis*, *A. ibitiensis*, *A. piracicabae*, *Apareiodon* sp. (Rio Verde), *A. vladii*, *A. vittatus* e
30 *P. nasus* (JESUS; MOREIRA-FILHO, 2000a, b; JORGE; MOREIRA-FILHO, 2004; ROSA
31 et al., 2006; VICARI et al., 2006b; BELLAFRONTTE et al., 2005, 2009, 2011; AZAMBUJA

1 et al., 2022b). Em *A. argenteus*, *A. machrisi*, *Apareiodon* sp. 2 (Córrego Bandeirinha), *P.*
2 *hilarii*, *P. moreirai*, *P. pongoensis* e *S. wagneri* o rDNA 18S encontra-se localizado na
3 região terminal de um par metacêntrico/submetacêntrico (VICENTE; JESUS; MOREIRA-
4 FILHO, 2001; CENTOFANTE; BERTOLLO; MOREIRA-FILHO, 2002; TRALDI et al.,
5 2016, 2020; NIRCHIO et al., 2021). Em *A. cavalcante*, *A. davisii*, *A. hasemani* e
6 *Apareiodon* sp. (Rio Aripuanã), múltiplos sítios de rDNA 18S foram observados
7 (BELLAFRONTTE et al., 2012; TRALDI et al., 2016, 2020; SANTOS et al., 2019). Sítios
8 adicionais de 18S ainda foram descritos em indivíduos de *A. ibitiensis* e *A. vittatus*
9 (BELLAFRONTTE et al., 2009, 2011).

10 Já para o rDNA 5S, a localização deste rDNA próxima à região centromérica de
11 um cromossomo metacêntrico/submetacêntrico é descrito nas espécies *A. affinis*, *A.*
12 *argenteus*, *A. ibitiensis*, *A. piracicabae*, *Apareiodon* sp. (Rio Verde), *Apareiodon* sp. (Rio
13 Aripuanã), *Apareiodon* sp. 2 (Córrego Bandeirinha), *A. vittatus*, *P. hilarii*, *P. moreirai*, *P.*
14 *pongoensis* e *S. wagneri* (CENTOFANTE; BERTOLLO; MOREIRA-FILHO, 2002; VICARI
15 et al., 2006b; BELLAFRONTTE et al., 2009, 2011; SANTOS et al., 2019; NIRCHIO et al.,
16 2021). Em *P. nasus*, este encontra-se localizado na região proximal de um cromossomo
17 subtelocêntrico (BELLAFRONTTE et al., 2005; AZAMBUJA et al., 2022b), e em *A.*
18 *hasemani* na região terminal de um par submetacêntrico (BELLAFRONTTE et al., 2012).
19 Nas espécies *A. cavalcante*, *A. davisii*, *A. machrisi* e *A. vladii* sítios múltiplos foram
20 relatados (ROSA et al., 2006; TRALDI et al., 2016, 2020).

21 *Parodon nasus* e *S. wagneri* apresentam os clusters de rDNAs em sintonia,
22 enquanto nas espécies *A. cavalcante*, *A. davisii*, *A. machrisi* estes encontram-se
23 colocalizados em alguns pares cromossômicos (BELLAFRONTTE et al., 2005; TRALDI et
24 al., 2016, 2020; NIRCHIO et al., 2021; AZAMBUJA et al., 2022b).

25 Genes histônicos foram mapeados em apenas seis espécies de *Apareiodon* e
26 em uma espécie de *Parodon* por Traldi et al. (2019). Para a histona H1, que possui
27 similaridade ao TE ERV1-2_FCa-I_1, um bloco na região proximal de um par
28 submetacêntrico foi observado, em adição à sinais dispersos por todos os cromossomos
29 do complemento. A histona H4, foi mapeada colocalizada ao bloco observado para a
30 histona H1 (TRALDI et al., 2019).

1 Espécies de Parodontidae são conhecidas por apresentarem diferentes sistemas
2 de cromossomos sexuais, com a ocorrência de espécies sem cromossomos sexuais
3 heteromórficos, com proto-cromossomos sexuais, sistema de cromossomos sexuais
4 simples (ZZ/ZW) ou múltiplo (ZZ/ZW₁W₂) (SCHEMBERGER et al., 2011).

5 Análise das regiões heterocromáticas, e o mapeamento das sequências
6 repetitivas satélite pPh2004 (VICENTE et al., 2003) e WAp (VICARI et al., 2010) ajudaram
7 a elucidar a origem dos diferentes sistemas de cromossomos na família
8 (SCHEMBERGER et al., 2011). Para o sistema ZZ/ZW, a origem é proposta ter ocorrido
9 a partir de uma inversão paracêntrica de um sítio terminal portador da fração denominada
10 WAp para a região proximal do braço curto de um par metacêntrico, com posterior
11 amplificação desta região levando a diferenciação do cromossomo W nas espécies
12 portadoras deste sistema (SCHEMBERGER et al., 2011), com exceção de *A. hasemani*,
13 onde há um pequeno acúmulo de heterocromatina e sequências WAp no braço curto do
14 cromossomo W (BELLAFRONTTE et al., 2012).

15 Para a origem do sistema múltiplo, é proposto inicialmente uma translocação
16 recíproca entre um par autossômico e um par de proto cromossomo sexual, depois o
17 cromossomo W sofreu uma fissão cêntrica seguida por uma inversão pericêntrica para
18 dar origem aos cromossomos W₁ e W₂ (MOREIRA-FILHO; BERTOLLO; GALETTI Jr,
19 1980; SCHEMBERGER et al., 2011).

20 Sequências do satélite de DNA pPh2004 foram mapeadas nos cromossomos
21 sexuais de *P. hilarii*, *P. moreirai*, *A. affinis* e *Apareiodon* sp. (Rio Aripuanã) e contribuíram
22 para a diferenciação destes (VICENTE et al., 2003; SCHEMBERGER et al., 2011;
23 NASCIMENTO et al., 2018; SANTOS et al., 2019). Neste contexto, na família são
24 descritas espécies sem cromossomos sexuais heteromórficos (*A. piracicabae* e *A.*
25 *vittatus*); com proto cromossomos sexuais e presença de DNA satélite pPh2004 (*Parodon*
26 *nasus*, *P. pongoensis*, *A. argenteus*, *A. cavalcante*, *A. davisii*, *A. machrisi* e *Apareiodon*
27 sp. 2 – Córrego Bandeirinha); com sistema de cromossomos sexuais heteromórficos
28 (ZZ/ZW) (*Apareiodon* sp. – Rio Verde, *A. ibitiensis*, *A. vladii* e *A. hasemani*); com sistema
29 de cromossomos sexuais heteromórficos e presença de DNA satélite pPh2004
30 (*Apareiodon* sp. – Rio Aripuanã, *P. moreirai* e *P. hilarii*); além da espécie *A. affinis*
31 (sistema do Alto Rio Paraná), com sistema de cromossomos sexuais múltiplos

1 (ZZ/ZW₁W₂) e DNA satélite pPh2004 (SCHEMBERGER et al., 2011; BELLAFRONTE et
2 al., 2012; TRALDI et al., 2016, 2020; NASCIMENTO et al., 2018; SANTOS et al., 2019).
3 Recentemente, a primeira caracterização citogenética para o gênero *Saccodon* foi
4 descrita por Nirchio et al. (2021), demonstrando que *S. wagneri* possui sistema de
5 cromossomos sexuais heteromórficos ZZ/ZW.

6 Estudos utilizando microssatélites e TEs, demonstram ainda o envolvimento
7 destas sequências na diferenciação dos cromossomos sexuais na família
8 (SCHEMBERGER et al., 2014, 2016, 2019; ZIEMNICZAK et al., 2014). Recentemente o
9 genoma de macho e fêmea de *Apareiodon* sp. (Rio Verde) foi montado (SCHEMBERGER
10 et al., 2019). Análises de anotação da fração repetitiva do genoma demonstram que cerca
11 de 36% deste correspondem a elementos repetitivos, como microssatélites, satélites de
12 DNA e TEs (SCHEMBERGER et al., 2019).

1 2 OBJETIVOS

2

3 2.1 OBJETIVO GERAL

4

5 Avaliar a ocorrência de ncRNA derivados de miniaturas de elementos
6 transponíveis (MITEs) e a organização de famílias gênicas U snRNA no genoma de
7 *Apareiodon* sp. com enfoque no reconhecimento de moléculas cooptadas, bem como, na
8 compreensão da atuação desses elementos de repetição na diferenciação cromossômica
9 em espécies de Parodontidae.

10

11 2.2 OBJETIVOS ESPECÍFICOS

12

- 13 • Prospectar e caracterizar sequências de miniaturas de elementos transponíveis
- 14 (MITEs) no genoma de *Apareiodon* sp.;
- 15 • Realizar a identificação de ncRNAs derivados a partir das sequências MITEs;
- 16 • Prospectar e caracterizar genes da família multigênica dos U snRNAs no genoma
- 17 de *Apareiodon* sp.;
- 18 • Investigar a participação de sequências de U snDNAs na diversificação dos
- 19 cromossomos da família Parodontidae.

1 3 MATERIAL E MÉTODOS

2

3 3.1 MATERIAL BIOLÓGICO

4

5 Exemplares da família Parodontidae foram coletados nas bacias Alto Rio Paraná,
6 Baixo Rio Paraná, São Francisco, Tocantins-Araguaia e Amazônica (Tabela 1, Figura 1).
7 As coletas foram realizadas com autorização do Instituto Chico Mendes de Conservação
8 da Biodiversidade (SISBIO: 15117-3) e Sistema Nacional de Gestão do Patrimônio
9 Genético e do Conhecimento Tradicional Associado (SISGEN: AE12D3D). Os
10 procedimentos foram aprovados pela Comissão de Ética no Uso Animal da Universidade
11 Estadual de Ponta Grossa (Processo CEUA 06/2019) e pela Comissão Técnica Nacional
12 de Biossegurança – CTNBio (CQB: 0063/98).

13 Os peixes coletados foram mantidos em aquários aerados e utilizados para
14 obtenção de cromossomos mitóticos a partir de células renais pelo método *air-drying*
15 (BERTOLLO; CIOFFI; MOREIRA-FILHO, 2015). Indivíduos de cada população analisada
16 foram depositados nas coleções ictiológicas do Núcleo de Pesquisas em Limnologia,
17 Ictiologia e Aquacultura (NUPELIA) da Universidade Estadual de Maringá e do Museu
18 Nacional do Rio de Janeiro (MNRJ). Os indivíduos de *A. piracicabae* do rio Mogi-Guaçu
19 e *P. hilarii* do ribeirão Araras encontram-se no Laboratório de Biologia Cromossômica:
20 Estrutura & Função (CBSFLab) da Universidade Estadual de Ponta Grossa e serão
21 encaminhados para depósito nos museus.

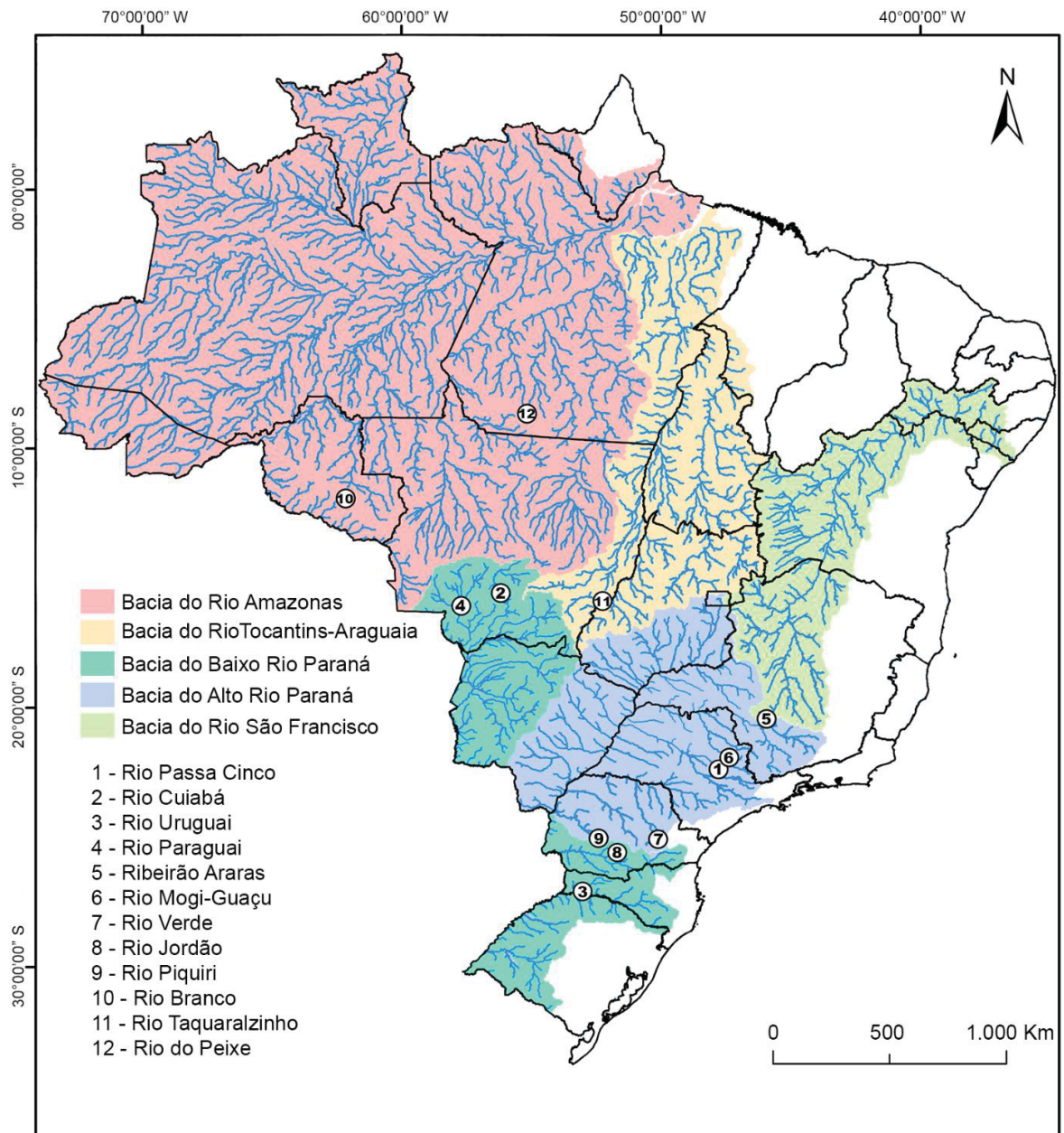
22

23 **Tabela 1** – Espécies da família Parodontidae estudadas, localização, número de indivíduos utilizados e
24 código de depósito.

Espécie	Rio/Bacia hidrográfica	Número		
		de indivíduos	Voucher	GPS
<i>Apareiodon affinis</i>	Rio Passa Cinco/Bacia do Alto Rio Paraná	12♂ 14♀	NUP16983	22°22'19" S; 47°46'54" W
<i>Apareiodon affinis</i>	Rio Cuiabá/Bacia do Baixo Rio Paraná	5♂ 6♀	NUP16265	15°34'41" S; 56°09'59" W
<i>Apareiodon affinis</i>	Rio Uruguai/ Bacia do Baixo Rio Paraná	8♂ 9♀	NUP16270	27°05'21" S; 53°01'01" W

Espécie	Rio/Bacia hidrográfica	Número de indivíduos	Voucher	GPS
<i>Apareiodon affinis</i>	Rio Paraguai/Bacia do Baixo Rio Paraná	7 ♂ 9 ♀	NUP16264	16°04'31" S; 57°42'10" W
<i>Apareiodon ibitiensis</i>	Ribeirão Araras/Bacia do Rio São Francisco	8 ♂ 9 ♀	MNRJ32741	20°26'16" S; 45°55'39" W
<i>Apareiodon ibitiensis</i>	Rio Passa Cinco/Bacia do Alto Rio Paraná	5 ♂ 7 ♀	MNRJ32771	22°22'19" S; 47°46'54" W
<i>Apareiodon piracicabae</i>	Rio Mogi-Guaçu/Bacia do Alto Rio Paraná	8 ♂ 10 ♀	-	21°55'34" S; 47°22'03" W
<i>Apareiodon sp.</i>	Rio Verde/Bacia do Alto Rio Paraná	15 ♂ 17 ♀	NUP3447	25°05'02" S; 50°05'50" W
<i>Apareiodon vittatus</i>	Rio Jordão/Bacia do Baixo Rio Paraná	5 ♂ 3 ♀	NUP19987	25°35'50.3" S; 51°40'44.6" W
<i>Apareiodon vladii</i>	Rio Piquiri/Bacia do Alto Rio Paraná	10 ♂ 10 ♀	NUP3375; NUP3376	25°02'57.6" S; 52°24'22.7" W
<i>Parodon buckleyi</i>	Rio Branco/Bacia do Rio Amazonas	22 ♂ 10 ♀	NUP23214	11°55'51.1" S; 62°09'09.4" W
<i>Parodon hilarii</i>	Ribeirão Araras/Bacia do Rio São Francisco	9 ♂ 14 ♀	-	20°26'16" S; 45°55'39" W
<i>Parodon nasus</i>	Rio Passa Cinco/Bacia do Alto Rio Paraná	7 ♂ 9 ♀	NUP23219	22°22'19" S; 47°46'54" W
<i>Parodon pongoensis</i>	Rio Taquaralzinho/Bacia do Tocantins-Araguaia	4 ♂ 5 ♀	NUP12147	15°53'28" S; 52°14'56" W
<i>Parodon sp.</i>	Rio do Peixe/Bacia do Rio Amazonas	8 ♂ 2 ♀	NUP23216	08°39'15.5" S; 55°09'24.3" W

- 1 **Figura 1** – Mapa do Brasil com a localização dos pontos de coleta dos espécimes de Parodontidae
 2 analisados.



3

1 3.2 MÉTODOS

2

3 3.2.1 Sequenciamento e montagem do genoma de *Apareiodon* sp.

4

5 Duas montagens do genoma de *Apareiodon* sp. foram utilizadas neste trabalho.
6 Para o capítulo 1, uma nova montagem utilizando *longs reads* e *shorts reads* foi realizada,
7 enquanto para o capítulo 2, a versão do genoma de fêmea publicada por Schemberger
8 et al. (2019) e disponível em [http://sacibase.ibb.unesp.br/jbrowse/JBrowse-](http://sacibase.ibb.unesp.br/jbrowse/JBrowse-1.12.1/index.html?data=data-apareiodon-sp)
9 [1.12.1/index.html?data=data-apareiodon-sp](http://sacibase.ibb.unesp.br/jbrowse/JBrowse-1.12.1/index.html?data=data-apareiodon-sp) foi utilizada.

10 Para a montagem da nova versão do genoma de *Apareiodon* sp., inicialmente
11 *Longs reads* foram obtidos a partir do isolamento de DNA de músculo de uma fêmea de
12 *Apareiodon* sp. (Rio Verde) na empresa Novagen Technology Inc (Califórnia, EUA). Após
13 avaliação da qualidade da extração do DNA, este foi utilizado para construção da
14 biblioteca e sequenciamento na plataforma PacBio. A qualidade das bibliotecas de *short*
15 *reads* (Illumina Hiseq 100 *base paired-end reads* – macho e fêmea; Illumina Hiseq 150
16 *base paired-end reads* – macho e fêmea) obtidas por Schemberger et al. (2019) foram
17 verificadas no FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), e
18 filtradas com a ferramenta Trimmomatic (BOLGER; LOHSE; USADEL, 2014) para
19 remoção dos *reads* de baixa qualidade (Q score < 33) e captura de *reads* >80 ou >140
20 nucleotídeos para cada uma das bibliotecas.

21 Primeiramente as *long reads* obtidas foram submetidas à montagem genômica
22 no Canu (KOREN et al., 2017, 2018; NURK et al., 2020) utilizando os seguintes
23 parâmetros: CorrectedErrorRate=0,035, utgOvlErrorRate=0,065, trimReadsCoverage=2,
24 trimReadsOverlap=500. Uma versão do genoma haploide foi removida usando o software
25 Redudans (PRYSZCZ; GABALDÓN, 2016). Para correção e polimento da montagem,
26 foram realizadas 3 rodadas no Pilon (WALKER et al., 2014), utilizando a biblioteca de
27 *short reads* de *Apareiodon* sp. Após, dois procedimentos de obtenção de *scaffolds* foram
28 realizados utilizando o RaGOO (ALONGE et al., 2019). No primeiro, os genomas de
29 referência de *Pygocentrus nattereri* (piranha de barriga vermelha) (NCBI RefSeq Access:
30 GCF_015220715.1), *Carassius auratus* (goldfish) (NCBI RefSeq Access:
31 GCF_003368295.1), e *Astyanax mexicanus* (tetra mexicano) (NCBI RefSeq Access:

1 GCF_000372685.2) foram utilizados. Em seguida, o segundo procedimento de
2 *scaffolding* foi realizado utilizando apenas o genoma de referência de *P. nattereri*,
3 removendo as sequências NC_051239.1 e NC_051240.1 (menores cromossomos do
4 genoma – prováveis cromossomos sexuais), seguido por uma rodada final no Pilon.

5 A verificação da qualidade da montagem foi realizada a partir da busca de genes
6 ortólogos no BUSCO (SIMÃO et al., 2015) no software gVolante (NISHIMURA; HARA;
7 KURAKU, 2017).

8 A anotação genômica foi realizada com o *pipeline* BRAKER2 (BRUNA et al.,
9 2021), usando como referência o conjunto completo de proteínas de *P. nattereri*, *C.*
10 *auratus* e *A. mexicanus*.

11

12 3.2.2 Prospecção e caracterização de *Miniature Inverted-repeats Transposable Elements*
13 no genoma de *Apareiodon* sp.

14

15 A prospecção das MITEs foi realizada no genoma de *Apareiodon* sp. (versão *long*
16 *reads*) com a utilização da ferramenta MITE tracker (CRESCENTE et al., 2018) com os
17 parâmetros padrões. Inicialmente, MITEs candidatos foram identificados a partir da busca
18 de sequências com repetições invertidas válidas com tamanho entre 50 e 800
19 nucleotídeos. Uma busca por BLAST (*nucleotide – nucleotide*) foi utilizada para alinhar
20 cada candidato à sua sequência reverso complemento. Após a identificação das regiões
21 invertidas terminais (TIR) válidas, a existência de *target site duplications* (TSDs) foram
22 checadas nas regiões à esquerda e direita das TIRs. As MITEs válidas (presença de TIR
23 e TSD) foram alinhadas e agrupadas em famílias utilizando o VSEARCH (ROGNES et
24 al., 2016) considerando as sequências TIR e interna, e similaridade de 80%. Sequência
25 TSDs não foram consideradas nesta etapa, uma vez que podem mudar durante as
26 inúmeras transposições das sequências. Para cada elemento identificado, cerca de 50
27 nucleotídeos à esquerda e a direita da TSD foram selecionados e comparados às regiões
28 flangeadoras de cada elemento da família. Apenas elementos que diferem
29 completamente nas suas regiões flangeadoras foram consideradas cópias únicas
30 pertencentes a mesma família. Após a comparação de todas as cópias da família, esta
31 foi considerada conservada se apresentasse ao menos 3 cópias distintas.

1 Para classificação das famílias de MITEs identificadas, as sequências obtidas
2 foram submetidas à busca por similaridade no CENSOR (KOHANY et al., 2006) em
3 adição a análise das TIRs.

4 O mapeamento das MITEs prospectadas no genoma de *Apareiodon* sp., em um
5 contexto gênico, foi realizado a partir da utilização de um arquivo GFF3 contendo as
6 coordenadas das MITEs e os arquivos GFF3 de anotação gênica gerados pelo
7 BRAKER2. Inicialmente, as posições das MITEs foram comparadas às regiões exônicas
8 e intrônicas anotadas no genoma de *Apareiodon* sp. utilizando a opção “intersect” na
9 ferramenta BEDtools (QUILAN; HALL, 2010). As MITEs identificadas nestas duas regiões
10 foram removidas do arquivo GFF3, e este foi comparado às coordenadas das regiões
11 gênicas utilizando a opção “closest” na ferramenta BEDtools para determinar a distância
12 ao gene mais próximo. De acordo com a localização das sequências, essas foram
13 classificadas em: intrônicas (IN – MITE localizada dentro de intron), exônicas (EX – MITE
14 localizada dentro de éxon), *upstream* (UP – MITE localizada até 10,000 nt da região 5’
15 do gene), *downstream* (DW – MITE localizada até 10,000 nt da região 3’ do gene), e
16 intergene (IG) quando nenhum dos critérios anteriores foi observado.

17 18 3.2.3 Identificação de ncRNAs nas MITEs

19
20 Buscas por ncRNAs foram realizadas nas MITEs prospectadas no genoma de
21 *Apareiodon* sp. Inicialmente, uma busca por famílias de ncRNAs conservadas foi
22 realizada com a ferramenta Infernal 1.1.3 (NAWROCKI; EDDY, 2013) e os modelos de
23 covariância disponíveis no Rfam (v. 14.8 – maio 2022) (KALVARI et al., 2020). Os
24 ncRNAs preditos foram filtrados considerando apenas as interações válidas e um E-value
25 $\leq 1.0e-5$. A sequência consenso para cada família identificada foi alinhada ao MITE para
26 identificar regiões conservadas utilizando CLUSTALW no Geneious v.7.1.9 (KEARSE et
27 al., 2012).

28 Em um segundo momento, um banco de dados de miRNAs maduros de peixes
29 foi obtido a partir do *release* 22.1 do miRBase (KOZOMARA; BIRGAOANU; GRIFFITHS-
30 JONES, 2019). Sequências para as espécies *Astatotilapia burtoni* (abu-mir – 236
31 sequências), *Cyprinus carpio* (ccr-mir – 146 sequências), *Danio rerio* (dre-mir – 374

1 sequências), *Electrophorus electricus* (eel-mir – 34 sequências), *Fugu rubripes* (fru-mir –
2 108 sequências), *Gadus morhua* (gmo-mir – 516 sequências), *Hippoglossus*
3 *hippoglossus* (hhi-mir – 36 sequências), *Ictalurus punctatus* (ipu-mir – 205 sequências),
4 *Metriaclima zebra* (mze-mir – 184 sequências), *Neolamprologus brichardi* (nbr-mir – 182
5 sequências), *Oryzias latipes* (ola-mir – 146 sequências), *Oreochromis niloticus* (oni-mir –
6 695 sequências), *Pundamilia nyererei* (pny-mir – 182 sequências), *Paralichthys olivaceus*
7 (pol-mir – 38 sequências), *Salmo salar* (ssa-mir – 497 sequências), e *Tetraodon*
8 *nigroviridis* (tni-mir – 109 sequências) foram recuperadas. Para evitar sequências
9 redundantes, sequências idênticas foram agrupadas utilizando o software CD-HIT
10 (<http://weizhong-lab.ucsd.edu/cd-hit/>) e filtradas do banco de dados. A busca por miRNAs
11 nas sequências MITEs foi realizada a partir de BLASTn local, utilizando os parâmetros -
12 task blastn-short e -word_size 15.

13

14 3.2.4 Predição e caracterização de snRNAs em *Apareiodon* sp.

15

16 snRNAs que compõem o *major* e o *minor* spliceossomos foram preditos no
17 genoma de *Apareiodon* sp. (*shorts reads version* – SCHEMBERGER et al., 2019)
18 utilizando o software Infernal 1.1.3 (NAWROCKI; EDDY, 2013), empregando as
19 bibliotecas dos modelos de covariância disponíveis no Rfam (KALVARI et al., 2020) (v.
20 14.2 – abril 2020) (U1 – RF00003; U2 – RF00004; U4 – RF00015; U5 – RF00020; U6 –
21 RF00026; U11 – RF00548; U12 – RF00007; U4atac – RF00618; U6atac – RF00619) de
22 acordo com o Protocolo Alternativo 1 (KALVARI et al., 2018).

23

24 Para a predição da estrutura secundária, as 10 sequências com menores valores
25 de E-value para cada família dos snRNAs *U1*, *U2*, *U4*, *U5* e *U6* foram alinhadas utilizando
26 CLUSTALW no Geneious v.7.1.9 (KEARSE et al., 2012). Já para os snRNAs *U11*, *U12*,
27 *U4atac* e *U6atac* todas as sequências preditas foram alinhadas. A ferramenta RNAalifold
28 (BERNHART et al., 2008), disponível no pacote ViennaRNA 2.0 (LORENZ et al., 2011),
29 foi utilizada para obtenção das estruturas secundárias. A formação de hétero-dímeros
30 entre os snRNAs *U4/U6* e *U4atac/U6atac* foi verificada a partir das sequências consenso
31 destes snRNAs, utilizando a ferramenta RNAcofold para determinar a interação
(BERNHART et al., 2006).

1 3.2.5 Identificação e caracterização de regiões promotoras dos snRNAs

2

3 O software MEME suite v. 5.3.3 (BAILEY et al., 2009), que inclui a ferramenta
4 MEME (Número do *motif* por sequência = 0 ou 1; número máximo de *motifs* = 10; tamanho
5 dos *motifs* = 5 – 25; número de nucleotídeos nos *motifs* = 5 – 20) (BAILEY et al., 2009)
6 foi utilizado para a descoberta de *motifs* conservados até 200 pb na região *upstream* dos
7 snRNAs. Após a identificação dos *motifs*, a busca pelos elementos conservados *Proximal*
8 *Sequence Element* (PSE) e TATA-box foi realizada manualmente nas sequências dos
9 snRNAs.

10

11 3.2.6 Desenho dos *primers* e confirmação dos snRNAs no genoma de *Apareiodon* sp.

12

13 *Primers* para cada família de snRNA foram desenhados a partir do alinhamento
14 das sequências preditas no genoma de *Apareiodon* sp. com menores E-values no
15 software *online* Primer3Plus (UNTERGASSER et al., 2007). O DNA genômico de
16 *Apareiodon* sp. foi extraído do fígado pelo método CTAB (cetyltrimethylammonium)
17 (MURRAY; THOMPSON, 1980) e utilizado como molde para Reações em Cadeia da
18 Polimerase (PCR), para obtenção de amplicons de cada snRNA. As reações de PCR
19 continham: 1X tampão de reação *Taq* DNA polimerase (200 mM Tris pH 8.4, 500 mM
20 KCl) (Invitrogen®), 2.5 mM de MgCl₂ (Invitrogen®), 0.2 mM de dNTPs, 0.5 µM de cada
21 *primer*, 1 U de *Taq* DNA polimerase (Invitrogen®) e 40 ng de DNA. O programa de reação
22 consistiu de desnaturação inicial por 10 minutos à 95°C, seguido de 35 ciclos de 95°C
23 por 1 minuto, 61,5°C por 45 segundos e 72°C por 1 minuto, e extensão final por 10
24 minutos à 72°C. Os produtos de PCR obtidos foram purificados utilizando o kit Illustra
25 GFX PCR DNA and Gel Band Purification (GE Healthcare), clonados em vetor pGEM
26 (Promega®) e transformados na bactéria *Escherichia coli* DH5α (Gibco GBRL®). O
27 isolamento e extração do DNA plasmidial foi realizado pelo método de lise alcalina, e
28 sequenciados em sequenciador automático ABI-prism 3500 Genetic Analyzer (Applied
29 biosystems). As sequências de DNA obtidas foram corrigidas no *software* Geneious
30 v.7.1.9 (KEARSE et al., 2012) e submetidas aos bancos de dados Basic Local Alignment
31 Search Tool (BLASTn) (ALTSCHUL et al., 1990), Rfam (KALVARI et al., 2020) e

1 CENSOR (GIRINST) (KOHANY et al., 2006) para caracterização destas. Todas as
2 sequências obtidas foram depositadas no GenBank.

3

4 3.2.7 Hibridação *in situ* Fluorescente (FISH)

5

6 As sequências clonadas e confirmadas como snRNAs U1, U2, U4, U5 e U6 foram
7 amplificadas por PCR como sondas utilizando os nucleotídeos marcados digoxigenina-
8 11dUTP (Jena Bioscience, Jena, Germany) (U1, U2 e U4), Aminoallyl-dUTP-Cy5 (Jena
9 Bioscience, Jena, Germany) (U2 e U4), e biotina-16-dUTP (Jena Bioscience, Germany)
10 (U5 e U6), e usados nos experimentos de localização *in situ*. A Hibridação *in situ*
11 Fluorescente (FISH) foi realizada de acordo com o protocolo geral descrito por PINKEL;
12 STRAUME; GRAY (1986), e a etapa de hibridação foi realizada com as seguintes
13 condições de estringência: 200ng de cada sonda, 50% formamida, 10% sulfato de
14 dextrano, e 2xSSS – Solução salina de citrato de sódio, por 16 horas à 37 °C em câmara
15 úmida.

16 As sondas marcadas com biotina-16dUTP foram detectadas com Streptavidin
17 Alexa Fluor 488 (Molecular Probes, Eugene, Germany) e as sondas marcadas com
18 digoxigenina-11-dUTP com anti-digoxigenin rhodamine Fab fragments (Roche Applied
19 Science, Penzberg, Germany), e os cromossomos foram contracolorados com solução
20 DAPI (0.2 µL/mL) montado em meio VectaShield (Vector). As melhores metáfases foram
21 fotografadas em microscópio de epifluorescência Leica (DM 2000) equipado com câmera
22 CCD Leica DFC3000 G. Os cromossomos foram classificados em metacêntrico (m),
23 submetacêntrico (sm), subtelocêntrico (st) e acrocêntrico (a) de acordo com a regra da
24 razão dos braços (LEVAN, FREDGA; SANDBERG, 1964) e organizados em cariótipos.

1 **4 RESULTADOS**

2

3 Os resultados obtidos neste estudo estão apresentados em forma de artigos
4 científicos divididos em 3 capítulos:

5

6

7 4.1 CAPÍTULO 1

8

9 Miniaturas de elementos transponíveis no genoma de *Apareiodon* sp. (Characiformes:
10 Parodontidae) – Em elaboração.

11

12 4.2 CAPÍTULO 2

13

14 Major and minor U small nuclear RNAs genes characterization in a neotropical fish
15 genome: Chromosomal remodeling and repeats units dispersion in Parodontidae.
16 Publicado na revista Gene.

17

18 4.3 CAPÍTULO 3

19

20 Comparative U2 and U4 snDNA chromosomal mapping in the Neotropical fish genera
21 *Apareiodon* and *Parodon* (Characiformes: Parodontidae) – Em elaboração.

1 4.1 CAPÍTULO 1

2

3

4 Miniaturas de elementos transponíveis no genoma de *Apareiodon* sp. (Characiformes:
5 Parodontidae)

6

7

8 Artigo em preparação

1 **Miniaturas de elementos transponíveis no genoma de *Apareiodon* sp.**
2 **(Characiformes: Parodontidae)**

3

4 **Resumo**

5 Miniaturas de elementos transponíveis (MITEs) são pequenos transposons de DNA não
6 autônomos presentes em grande número de cópias em genomas eucariotos. Inserção de
7 MITEs em introns ou próximas à genes são comuns, e tendem a alterar a estrutura gênica
8 e sua expressão, fornecendo uma nova fonte de diversidade funcional. A origem de
9 miRNAs derivadas de MITEs foi descrita em inúmeros organismos em um mecanismo
10 denominado de co-opção molecular. No presente estudo, a prospecção das MITEs foi
11 realizada no genoma do peixe neotropical *Apareiodon* sp. a partir do software MITE
12 Tracker. 9.229 sequências de MITEs foram identificadas e agrupadas em 425 famílias. A
13 identificação das MITEs demonstrou que as superfamílias *Tc1-Mariner* e *hAT* são as
14 responsáveis por cerca de 65% da diversidade e frequência das famílias. 81% das MITEs
15 encontram-se inseridas próximas ou dentro de genes. A busca por ncRNAs derivados de
16 MITEs revelaram a presença de 14 famílias de precursores de miRNAs distribuídas em
17 11 sequências de MITEs. Já a análise de similaridade identificou 29 diferentes miRNAs
18 maduros de peixes localizados em 50 MITEs de *Apareiodon* sp. O grande número de
19 MITEs localizadas próximas à genes em *Apareiodon* sp. as qualificam para processos de
20 co-opção molecular. Da mesma forma, a identificação de famílias de pre-miRNAs e
21 miRNAs maduros em sequências MITEs de *Apareiodon* sp. corroboram a proposta de
22 cooptação desses elementos para função ncRNA.

23

24 **Palavras-chave:** MITEs; genoma de peixe; co-opção molecular; miRNAs.

1 **Introdução**

2

3 Elementos transponíveis (TEs – *Transposable elements*) são sequências de DNA
4 que possuem a capacidade de se replicarem e se moverem nos genomas (CHÉNAIS et
5 al., 2012; WELLS e FESCHOTTE, 2020). Os TEs estão presentes na maioria dos
6 genomas e representam uma fração variável entre os eucariotos (KIDWELL, 2002;
7 KONING et al., 2011; TOLLIS; BOISSINOT, 2012; SOTERO-CAIO et al., 2017), sendo
8 uma importante fonte de variação genética e novidades genômicas (WELLS;
9 FESCHOTTE, 2020).

10 De acordo com Wicker et al. (2007), TEs podem ser classificados em duas
11 classes: Classe I ou Retrotransposons, os quais utilizam um intermediário de RNA para
12 se moverem no genoma, geralmente conhecidos como elementos transponíveis do tipo
13 “copia e cola”; e Classe II ou transposons de DNA, que se movem de maneira direta no
14 genoma sem a participação de um intermediário de RNA, a maioria deles conhecidos
15 como elementos “corta e cola”. Estes ainda são classificados em subclasses de acordo
16 com a mobilidade durante a transcrição reversa, e o número de fitas de DNA cortadas no
17 sítio doador (WICKER et al., 2007). Subsequentemente são agrupados em superfamílias,
18 baseado em características da estrutura proteica e domínios não codificantes, e famílias,
19 de acordo com a homologia e conservação da sequência de DNA (WICKER et al., 2007).

20 Miniaturas de elementos transponíveis (MITEs – *Miniature Inverted-repeat*
21 *Transposable Elements*) são Transposons de DNA não-autônomos, ou seja, sem
22 capacidade de transposição por enzima própria, por vezes caracterizados pelo alto
23 número de cópias, e com tamanho relativamente pequeno (50 – 800 pb) (WICKER et al.,
24 2007; CRESCENTE et al., 2018). MITEs possuem regiões invertidas terminais (TIRs –
25 *Terminal Inverted Regions*) e podem apresentar ainda duas repetições flanqueadoras
26 curtas, chamadas sítios alvo de duplicação (TSD – *Target Site Duplication*), geradas
27 durante a transposição (WICKER et al., 2007; CRESCENTE et al., 2018). No entanto,
28 durante o processo de transposição de uma MITE, é pouco provável que suas TSD sejam
29 transpostas juntas com o elemento (CRESCENTE et al., 2018).

30 MITEs foram inicialmente descritas no genoma de milho (BUREAU e WESSLER,
31 1992) e posteriormente identificadas em diferentes grupos de organismos (LIU et al.,

1 2019; MIKKELSEN; WEIR, 2020; ALBUQUERQUE; EBERT; HAAG, 2020; GRACE;
2 CARR, 2020). No entanto, estudos de prospecção e identificação de MITEs em genomas
3 de peixes ainda são escassos (IZSVÁK et al., 1999; SCHEMBERGER et al., 2016).

4 Apesar de MITEs e TEs autônomos serem capazes de ser transpostos pela
5 mesma transposase, seus números de cópias variam drasticamente nos genomas
6 (CHEN et al., 2014a). Enquanto famílias de elementos autônomos apresentam um ou
7 poucos membros, uma família de MITEs pode ter de dezenas a centenas de cópias
8 (FATTASH et al., 2013). Durante o processo de transposição, MITEs frequentemente são
9 inseridas próximas ou dentro de genes, podendo impactar a expressão destes (KUANG
10 et al., 2009; FATTASH et al., 2013; MACKO-PODGÓRNI; MACHAJ; GRZEBELUS, 2021;
11 KLAI et al., 2020, 2022). Apesar da transcrição das MITEs não serem requeridas para a
12 transposição, a localização destas próximas à genes e/ou em introns, permite que sejam
13 frequentemente transcritas (FESCHOTTE; JIANG; WESSLER, 2002; FESCHOTTE,
14 2008). Devido à presença das TIRs, MITEs transcritas podem formar estruturas de
15 dsRNA (*double strand RNA*) em grampo (*stem loop*), as quais são semelhantes às
16 sequências de pré-miRNA e que podem ser processadas, pela maquinaria de RNAi,
17 dando origem à microRNAs (miRNAs) (PIRIYAPONGSA; JORDAN, 2008; CUI; YOU;
18 CHEN, 2017).

19 *Apareiodon* sp. (espécie não descrita na literatura e próxima de *Apareiodon*
20 *ibitiensis*) é uma espécie de peixe neotropical pertencente à família Parodontidae
21 (VICARI et al., 2006b). Schemberger et al. (2019) demonstraram que cerca de 36% do
22 genoma de *Apareiodon* sp. corresponde à elementos repetitivos, sendo que destes, cerca
23 de 38% são Transposons de DNA. Duas fases de invasões de transposons de DNA foram
24 observadas na espécie: na fase mais antiga os elementos *Helitron*, *Tc1-mariner* e *EnSpm*
25 apresentaram expansão no número de cópias, enquanto que na segunda fase os
26 elementos *hAT*, *Harbinger*, *Crypton* e *Tc1-Mariner* estiveram envolvidos
27 (SCHEMBERGER et al., 2019). Neste contexto, o objetivo deste estudo foi realizar a
28 prospecção e caracterização de sequências MITEs no genoma de *Apareiodon* sp.,
29 identificando possíveis processos de co-opção molecular para função ncRNA.

30

31

1 **Material e métodos**

2

3 *Sequenciamento genômico e anotação gênica*

4

5 *Long reads* foram obtidos a partir do isolamento de DNA de músculo de uma
6 fêmea de *Apareiodon* sp. (Rio Verde, Paraná 25°04'35" S; 50°04'03" W) na empresa
7 Novagen Technology Inc (Califórnia, EUA), e após avaliação da qualidade submetido à
8 construção da biblioteca e sequenciamento na plataforma PacBio. *Short reads* foram
9 previamente obtidos por Schemberger et al. (2019) para dois indivíduos machos e duas
10 fêmeas (Illumina HiSeq 100 *base paired-end reads* – macho e fêmea; Illumina HiSeq 150
11 *base paired-end reads* – macho e fêmea).

12 Os dados Illumina foram submetidos a análise de qualidade no FastQC
13 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e filtrados com a ferramenta
14 Trimmomatic (BOLGER; LOHSE; USADEL, 2014) para remoção de *reads* de baixa
15 qualidade (Q Score < 33) e manutenção de *reads* >80 ou >140 nucleotídeos, de acordo
16 com as bibliotecas acima mencionadas.

17 As *long reads* foram submetidas à montagem no Canu (KOREN et al., 2017,
18 2018; NURK et al., 2020) com os seguintes parâmetros: CorrectedErrorRate=0,035,
19 utgOvlErrorRate=0,065, trimReadsCoverage=2, trimReadsOverlap=500. O genoma
20 haplótipo foi removido usando o software Redudans (PRYSZCZ; GABALDÓN, 2016). A
21 correção e polimento da montagem foi realizada em 3 rodadas no Pilon (WALKER et al.,
22 2014) com a biblioteca de *short reads* de fêmea de *Apareiodon* sp. Após, dois
23 procedimentos de obtenção de *scaffolds* foram realizados com o RaGOO (ALONGE et
24 al., 2019). Inicialmente os genomas de referência de *Pygocentrus nattereri* (piranha de
25 barriga vermelha) (NCBI RefSeq Access: GCF_015220715.1), *Carassius auratus*
26 (goldfish) (NCBI RefSeq Access: GCF_003368295.1), e *Astyanax mexicanus* (tetra
27 mexicano) (NCBI RefSeq Access: GCF_000372685.2) foram utilizados. Em seguida, o
28 segundo procedimento de *scaffolding* foi realizado utilizando apenas o genoma de
29 referência de *P. nattereri*, seguido por uma rodada final no Pilon. Para verificar a
30 qualidade da montagem, uma busca por genes ortólogos foi realizada no BUSCO
31 (SIMÃO et al., 2015) no gVolante (NISHIMURA; HARA; KURAKU, 2017).

1 A anotação genômica foi realizada com o *pipeline* BRAKER2 (BRÛNA et al.,
2 2021), usando como referência o conjunto completo de proteínas de *P. nattereri*, *C.*
3 *auratus* e *A. mexicanus*.

4 5 *Prospecção e caracterização das MITEs*

6
7 O genoma de *Apareiodon* sp. foi submetido à busca de MITEs na ferramenta
8 MITE tracker (CRESCENTE et al., 2018) com os parâmetros padrões. Inicialmente, uma
9 busca por sequências com repetições invertidas válidas com tamanho entre 50 e 800
10 nucleotídeos (nt) foi realizada para identificar MITEs candidatas. Uma busca por BLAST
11 (*nucleotide – nucleotide*) foi utilizada para alinhar cada candidato à sua sequência reverso
12 complemento. Após identificar TIR válida, a existência de TSDs foram checadas nas
13 regiões à esquerda e direita das TIRs. MITEs consideradas válidas foram alinhadas e
14 agrupadas em famílias pelo VSEARCH (ROGNES et al., 2016) considerando as
15 sequências TIR e sua região interna, e similaridade de 80%. Sequências TSDs foram
16 removidas nesta etapa, uma vez que estas podem mudar durante as diferentes
17 inserções. Para cada elemento, cerca de 50 nucleotídeos à esquerda e a direita da TSD
18 foram selecionados e comparados às regiões flangeadoras de cada elemento da família.
19 Apenas elementos que diferem completamente nas regiões flangeadoras foram
20 considerados sequências únicas pertencente à mesma família. Após a comparação de
21 todos as cópias da família, esta foi considerada conservada apenas se 3 ou mais
22 sequências diferentes foram observadas.

23 Para classificação das famílias de MITEs identificadas, as sequências MITEs
24 foram submetidas à busca por similaridade no CENSOR (KOHANY et al., 2006) em
25 adição a análise das TIRs.

26 27 *Mapeamento das MITEs no genoma de Apareiodon sp.*

28
29 Um arquivo GFF3 contendo as coordenadas das MITEs foi utilizado para
30 determinar a posição das sequências MITEs em um contexto gênico. Inicialmente, as
31 posições das MITEs foram comparadas às regiões exônicas e intrônicas anotadas no

1 genoma de *Apareiodon* sp. utilizando a opção intersect na ferramenta BEDtools
2 (QUINLAN; HALL, 2010). Em um segundo momento, MITEs não identificadas nestas
3 regiões foram comparadas às coordenadas das regiões transcritas utilizando a opção
4 closest na ferramenta BEDtools para determinar a distância ao gene mais próximo. De
5 acordo com a localização das sequências essas foram classificadas em: intrônicas (IN –
6 MITE localizada dentro de íntron), exônicas (EX – MITE localizada dentro de éxon),
7 *upstream* (UP – MITE localizada até 10.000 nt da região 5' do gene), *downstream* (DW –
8 MITE localizada até 10.000 nt da região 3' do gene), e intergene (IG) quando nenhum
9 dos critérios anteriores foi observado.

10

11 *Identificação de ncRNAs nas MITEs*

12

13 As MITEs prospectadas no genoma de *Apareiodon* sp. foram submetidas à uma
14 busca por ncRNAs conservados utilizando a ferramenta Infernal 1.1.3 (NAWROCKI;
15 EDDY, 2013) e os modelos de covariância do Rfam (v. 14.8 – may 2022) (KALVARI et
16 al., 2020). Os ncRNAs preditos foram filtrados considerando o limite de inclusão da
17 interação e um E-value $\leq 1.0e-5$. A sequência consenso para cada família identificada foi
18 alinhada ao MITE para identificar regiões conservadas utilizando CLUSTALW no
19 Geneious v.7.1.9 (KEARSE et al., 2012).

20

21 Um banco de dados de miRNAs maduros de peixes foi obtido a partir do release
22 22.1 do miRBase (KOZOMARA; BIRGAOANU; GRIFFITHS-JONES, 2019). Sequências
23 para as espécies *Astatotilapia burtoni* (abu-mir – 236 sequências), *Cyprinus carpio* (ccr-
24 mir – 146 sequências), *Danio rerio* (dre-mir – 374 sequências), *Electrophorus electricus*
25 (eel-mir – 34 sequências), *Fugu rubripes* (fru-mir – 108 sequências), *Gadus morhua* (gmo-
26 mir – 516 sequências), *Hippoglossus hippoglossus* (hhi-mir – 36 sequências), *Ictalurus*
27 *punctatus* (ipu-mir – 205 sequências), *Metriaclima zebra* (mze-mir – 184 sequências),
28 *Neolamprologus brichardi* (nbr-mir – 182 sequências), *Oryzias latipes* (ola-mir – 146
29 sequências), *Oreochromis niloticus* (oni-mir – 695 sequências), *Pundamilia nyererei* (pny-
30 mir – 182 sequências), *Paralichthys olivaceus* (pol-mir – 38 sequências), *Salmo salar*
31 (ssa-mir – 497 sequências), e *Tetraodon nigroviridis* (tni-mir – 109 sequências) foram
recuperadas. Para evitar sequências redundantes, sequências idênticas foram

1 agrupadas utilizando o software CD-HIT (<http://weizhong-lab.ucsd.edu/cd-hit/>) e apenas
2 um representante de cada agrupamento foi inserido na análise. A busca por miRNAs nas
3 sequências MITEs foi realizada através de BLASTn local utilizando os parâmetros -task
4 blastn-short e -word_size 15.

5

6 **Resultados**

7

8 *Montagem genômica*

9

10 A montagem *long reads*, seguida pela correção com as *short reads*, gerou um
11 primeiro *draft* genômico de *Apareiodon* sp. com 1.304 *contigs*. Durante o processo de
12 *scaffolding* utilizando como referência o genoma das espécies *P. nattereri*, *C. auratus* e
13 *A. mexicanus* resultou em um *draft* genômico com 277 *scaffolds*. Após um novo processo
14 de *scaffolding* utilizando o genoma de *P. nattereri* como referência resultou em 92
15 *scaffolds* com 954.157.697 nt (~0.95 GB). A qualidade da montagem foi verificada no
16 programa BUSCO, o qual demonstrou a presença de 195 genes completos (83.69%), 3
17 parciais (1.29%), e 35 faltantes (15.02%). A anotação resultante do programa BRAKER2
18 encontrou 36.290 genes e 13.933 proteínas similares ao genoma de referência.

19

20 *Prospecção e caracterização das MITEs*

21

22 Foram identificadas 9.229 sequências de MITEs no genoma de *Apareiodon* sp.,
23 as quais foram agrupadas em 425 famílias (Tabela suplementar 1 – disponível em:
24 [https://drive.google.com/file/d/1qppoCCFZvk6gnTKTBv4DGBieKJ7rqSX2/view?usp=sha](https://drive.google.com/file/d/1qppoCCFZvk6gnTKTBv4DGBieKJ7rqSX2/view?usp=sharing)
25 [ring](https://drive.google.com/file/d/1qppoCCFZvk6gnTKTBv4DGBieKJ7rqSX2/view?usp=sharing)). As MITEs apresentaram entre 46 e 800 pb, com tamanho médio de 506 pb,
26 representando cerca de 0,5% do genoma de *Apareiodon* sp. (4.757.417 nt). O tamanho
27 das TIRs variou de 10 a 651 pb, com tamanho médio de 27 pb. O número de sequências
28 observado dentro das famílias variou de 3 a 726, com valor médio de 22 sequências por
29 família.

30

31 A identificação das sequências MITEs demonstrou que suas famílias estão
agrupadas em 10 superfamílias de Transposons de DNA (Tabela 1). As MITEs das

1 superfamílias *hAT* e *Tc1-Mariner* foram as mais numerosas em diversidade de famílias e
2 em número de cópias encontradas no genoma de *Apareiodon* sp., correspondendo a
3 cerca de 65% das sequências prospectadas (Tabela 1).

4 A localização das MITEs no genoma de *Apareiodon* sp. revelou que 4.625
5 encontram-se na região intrônica, 1.493 na região *downstream* da região gênica
6 transcrita, 1.400 na região *upstream* da região transcrita, e 22 na região exônica (Figura
7 1, Tabela suplementar 1). Ainda, 1.689 MITEs foram descritas em regiões intergênicas.

8

9 *ncRNAs* derivados de MITEs

10

11 A busca por famílias de *ncRNAs* conservadas nas sequências MITEs,
12 demonstrou a ocorrência de 14 famílias de pre-miRNAs distribuídas em 11 sequências
13 de MITEs (Tabela 2). Nas MITEs 457, 458, 459, 461, 462, 463, 464, 8.440, e 8.565
14 diferentes famílias de miRNAs foram preditas. O alinhamento destas famílias às MITEs
15 revelou que estas encontram-se sobrepostas na mesma região da MITE (Figura 2).

16 Um total de 3.688 miRNAs maduros de peixes foram obtidos a partir do banco de dados
17 do miRBase. Após a remoção das cópias redundantes, 1.653 diferentes miRNAs foram
18 utilizados para busca de miRNAs nas sequências MITEs de *Apareiodon* sp. Cinquenta
19 diferentes MITEs apresentaram similaridade com 29 diferentes miRNAs (Tabela 3). Trinta
20 e uma destas MITEs encontram-se em íntrons, 10 na região intergênica, 8 na região
21 *downstream* e 1 na região *upstream* (Tabela 3).

22

23 **Discussão**

24

25 *Caracterização das MITEs no genoma de Apareiodon* sp.

26

27 Em plantas, as MITEs representam grande frações dos genomas, podendo
28 chegar à mais de 200.000 cópias e a aproximadamente 10% do genoma em algumas
29 espécies (CHEN et al., 2014b). No outro extremo, em espécies animais, as MITEs podem
30 apresentar poucas cópias, como observado na ave rendalinho-do-Xingu (*Willisornis vidua*
31 *nigrigula*) com 38 elementos agrupados em nove famílias (MIKKELSEN; WEIR, 2020).

1 Contudo, o número de sequências e famílias de MITEs é uma característica intrínseca
2 de cada espécie, podendo variar drasticamente quando espécies relacionadas são
3 comparadas (LIU et al., 2019; MACKO-PODGÓRNI; MACHAJ; GRZEBELUS, 2021; KLAI
4 et al., 2022). Esse é o primeiro estudo de identificação de MITEs utilizando análises de
5 larga escala em genomas de peixes, demonstrando um grande número de cópias (quase
6 10.000 sequências de MITEs) correspondendo a 0,5% do genoma de *Apareiodon* sp.

7 O genoma de *Apareiodon* sp. apresenta grande diversidade de superfamílias de
8 TEs, sendo os transposons de DNA *Tc1-Mariner* e *hAT* os mais abundantes
9 (SCHEMBERGER et al., 2019). Na fase de senescência de um TE, os elementos
10 progressivamente acumulam mutações, inserções e deleções em suas sequências, até
11 perderem completamente suas identidades (FERNÁNDEZ-MEDINA et al., 2012),
12 gerando cópias neutras ou cooptadas nos genomas (FESCHOTTE, 2008). As MITEs, por
13 sua vez, podem dispersar nos genomas a partir de transposons de DNA em diferentes
14 vias (FESCHOTTE; JIANG; WESSLER, 2002; FATTASH et al., 2013). O mecanismo
15 natural de origem de uma MITE envolve a deleção das regiões internas de um elemento
16 autônomo durante seu mecanismo de transposição, originando a MITE de mesmo nome
17 (FESCHOTTE; JIANG; WESSLER, 2002; FATTASH et al., 2013). Essas cópias não
18 autônomas podem amplificar seu número de cópias na presença de seus elementos
19 autônomos ativos (JIANG et al., 2004; FERNÁNDEZ-MEDINA et al., 2012). Em outra via,
20 quando os elementos autônomos de uma família MITE são perdidos ou inativados em
21 um genoma, algumas MITEs podem ainda ser mobilizadas de forma cruzada por
22 elementos autônomos relacionados, devido à presença de *motifs* conservados que
23 permitem a transposição (FATTASH et al., 2013). Nesse estudo, as MITEs das
24 superfamílias *Tc1-Mariner* e *hAT* foram as mais diversas em famílias e numerosas em
25 cópias no genoma de *Apareiodon* sp., o que está de acordo com a abundância e proposta
26 de invasão desses elementos no genoma da espécie (SCHEMBERGER et al., 2019).

27

28 *Co-opção molecular de sequências MITEs*

29

30 A ampla maioria das cópias MITEs mapeadas no genoma de *Apareiodon* sp.
31 estão localizadas dentro de genes ou próximas a eles. Esse fato sugere que a

1 conservação dessas sequências MITEs possa ser correlacionada a eventos de co-opção
2 molecular, ou seja, onde essas sequências possam servir para outras funções no
3 genoma. Tem sido proposto que as MITEs localizadas próximas ou dentro de genes
4 podem estar envolvidas em mecanismos funcionais nos genomas, como por exemplo, no
5 controle da expressão gênica, alteração de pontos de splicing e geração de ncRNAs
6 (PIRIYAPONGSA; JORDAN, 2008; KUANG et al., 2009; LU et al., 2012; DRONGITIS et
7 al., 2019).

8 O conceito de domesticação, co-opção ou exaptação molecular, foi descrito para
9 caracterizar o processo pelo qual uma sequência de TE é cooptada para realizar uma
10 nova função e trazer benefícios para o genoma (MILLER et al., 1997). Quando inseridas
11 na região *upstream*, as MITEs podem gerar novos sítios de ligação de fatores de
12 transcrição (TESTORI et al., 2012; HÉNAFF et al., 2014; MORATA et al., 2018). Por outro
13 lado, a presença de MITEs próximas à genes podem suprimir a expressão destes, através
14 da metilação do DNA (ARIEL; MANAVELLA, 2021). O grande número de MITEs
15 localizadas nas regiões *upstream* e *downstream* de genes no genoma de *Apareiodon* sp.
16 sugere que algumas destas sequências possam ter sido cooptadas, e estejam atuando
17 na regulação da expressão gênica.

18 Já a localização das MITEs em regiões intrônicas é indicativa da transcrição em
19 pré-RNA durante os eventos da expressão gênica. MITEs de localização intrônica podem
20 gerar sítios críticos de splicing e alternativas para a recomposição de éxons, ou mesmo,
21 atuar em alterações nos perfis de expressão, caso os introns possuam sequências
22 regulatórias (GREENE; WALKO; HAKE, 1994; KLAI et al., 2022). No genoma de
23 *Apareiodon* sp. cerca de 50% das MITEs identificadas tiveram localização intrônica, onde
24 a manutenção de suas sequências sugere a ocorrência de cópias cooptadas ou que
25 essas são neutras nos genomas e oriundas de mecanismo recente de dispersão.

26 Em outra via, a co-opção de cópias MITEs para origem de miRNAs foi proposta
27 inicialmente por Piriyaopngsa e Jordan (2008), e representa uma importante fonte de
28 origem destes ncRNAs nos genomas eucariotos (ZHANG; JIANG; GAO, 2011; CUI; YOU;
29 CHEN, 2017; ZHANG et al., 2018). miRNAs compõem uma classe de RNAs regulatórios
30 endógenos, com 19 a 25 nucleotídeos (LU; ROTHENBERG, 2018). miRNAs maduros são
31 gerados a partir de longos RNAs precursores contendo *stem-loop* (BORCHERT; LANIER;

1 DAVIDSON, 2006). *Stem-loop* formados em RNAs derivados de MITEs podem ser
2 substrato para a maquinaria de processamento de miRNAs (PIRIYAPONGSA; JORDAN,
3 2008). A atuação de um miRNA na regulação pós-transcricional envolve o
4 reconhecimento complementar de *trailers* de RNAm (geralmente região 3' UTR –
5 *untranslated region*) pelo *RNA-induced silencing complex* (VISHNOI; RANI, 2017).

6 Um grande número de miRNAs são evolutivamente conservados no reino animal
7 (SAETROM et al., 2006), o que possibilita a identificação de miRNAs em novas espécies
8 a partir de homologias (PAUL et al., 2018). A busca por famílias de ncRNAs na biblioteca
9 MITEs de *Apareiodon* sp. identificou 14 diferentes famílias precursoras de miRNAs.
10 Cinco dessas MITEs apresentaram identidade para uma única família miRNA, enquanto
11 as demais tiveram identidade para duas ou mais famílias decorrentes do alinhamento da
12 mesma sequência para miRNAs de diferentes organismos. A alta similaridade das
13 sequências para outros organismos reforça a proposição da região como um precursor
14 de miRNA, a qual pode ser processada pela maquinaria de RNAi e gerar miRNAs
15 maduros.

16 Entre as sequências MITEs de *Apareiodon* sp. com similaridade para precursores
17 de miRNA, sete encontram-se em regiões intrônicas ou próxima aos genes, as quais
18 podem ser transcritas a partir da expressão desses. Em outra via, cópias similares a
19 MITEs localizadas na região intergênica podem ter sido cooptadas como genes de
20 miRNAs, e apresentarem expressão independente. Em peixes neotropicais, apenas para
21 o peixe poraquê (*Electrophorus electricus*), sequências de miRNAs encontram-se
22 disponíveis em bancos de dados, fato que dificulta a busca de miRNAs em genomas por
23 similaridade. Por outro lado, comparando as MITEs de *Apareiodon* sp. com miRNAs
24 maduros descrito em 16 genomas de peixes foram identificados 29 diferentes miRNAs.
25 Os miRNAs *ola-miR-133-5p*, *dre-miR-203a-3p*, *dre-let-7c-3p* e *dre-miR-499-5p*
26 encontrados nas MITEs de *Apareiodon* sp. são descritas no genoma de diversos peixes,
27 demonstrando serem miRNAs conservados, enquanto os demais miRNAs anotados são
28 espécie-específicos. Apesar do baixo número de cópias MITEs carregando miRNAs,
29 muitas delas encontram-se localizadas em regiões intrônicas de genes, o que sugere a
30 sua transcrição. Os dados obtidos sugerem a co-opção de MITEs para a função miRNA
31 no genoma de *Apareiodon* sp., no entanto, dados de predição de alvos mRNA e formação

1 do *RNA-induced silencing complex* ainda são necessários, uma vez que um único miRNA
2 pode ter como alvo centenas de RNAm (LU; ROTHENBERG, 2018).

3

4 **Conclusão**

5

6 Este é o primeiro trabalho de prospecção e caracterização de MITEs utilizando
7 análises em larga escala nos genomas de peixes. A caracterização das MITEs
8 demonstrou que a maior parte destas sequências no genoma de *Apareiodon* sp. foram
9 originadas a partir de eventos de transposição de TEs pertencentes às superfamílias *Tc1-*
10 *Mariner* e *hAT*. A localização genômica das cópias MITEs próximas ou dentro dos genes
11 e a conservação de suas sequências nestas regiões sugere o seu envolvimento em
12 processos de co-opção molecular. Foi possível determinar inúmeros miRNAs nas MITEs
13 de *Apareiodon* sp. apesar da utilização de banco de dados de espécies não relacionadas
14 filogeneticamente. Contudo, análises mais robustas envolvendo o sequenciamento de
15 pequenos RNAs e de RNAs mensageiros permitirão a anotação de pre-miRNAs, miRNAs
16 maduros e pré-mRNAs. A partir desses será possível realizar a predição de alvos miRNAs
17 e a validação do processo de cooptação das MITEs para função ncRNA no genoma de
18 *Apareiodon* sp.

19

20 **Referências**

21

22 As referências deste capítulo encontram-se na seção referências ao final da tese.

1 **Tabela 1** – Classificação das sequências MITEs identificadas no genoma de *Apareiodon* sp.

Superfamília	Número de famílias	Número de sequências	Número médio de sequências por família	Número de cópias na família mais numerosa
<i>hAT</i>	152	2093	14	314
<i>Tc1-Mariner</i>	122	3825	31	726
<i>P-element</i>	36	1064	30	173
<i>PIF-Harbinger</i>	33	1227	37	269
<i>CACTA/EnSpm</i>	14	94	7	16
<i>Merlin</i>	13	55	4	8
<i>MuDR</i>	8	82	10	43
<i>PiggyBac</i>	4	22	6	11
<i>Dada</i>	3	123	41	62
<i>IS3EU</i>	2	25	13	17
<i>Unknown</i>	38	619	16	387
Total	425	9229	22	-

2

1 **Tabela 2** – Famílias de ncRNAs previstas nas sequências MITEs de *Apareiodon* sp.

MITE (localização genômica)	Família	ncRNA (Rfam)
MITE_T_457 (IG)	F24	mir-1277 microRNA precursor family (RF04154); mir-297 microRNA precursor family (RF04217); mir-3688 microRNA precursor family (RF04161); mir-466 microRNA precursor family (RF04052); MIR862 microRNA precursor family (RF04187)
MITE_T_458 (DW)	F24	mir-1277 microRNA precursor family (RF04154); mir-3149 microRNA precursor family (RF04176); mir-3688 microRNA precursor family (RF04161); mir-466 microRNA precursor family (RF04052); microRNA MIR1122 (RF00906)
MITE_T_459 (IN)	F24	mir-1277 microRNA precursor family (RF04154); mir-297 microRNA precursor family (RF04217); mir-3688 microRNA precursor family (RF04161); mir-4536 microRNA precursor family (RF03747); mir-466 microRNA precursor family (RF04052)
MITE_T_461 (DW)	F24	mir-1277 microRNA precursor family (RF04154); mir-297 microRNA precursor family (RF04217); mir-3149 microRNA precursor family (RF04176); mir-3688 microRNA precursor family (RF04161); mir-4536 microRNA precursor family (RF03747); mir-466 microRNA precursor family (RF04052); mir-4803 microRNA precursor family (RF03147); MIR2600 microRNA precursor family (RF03843)
MITE_T_462 (IG)	F24	mir-3149 microRNA precursor family (RF04176); mir-3688 microRNA precursor family (RF04161); mir-4536 microRNA precursor family (RF03747); mir-466 microRNA precursor family (RF04052); microRNA MIR1122 (RF00906)
MITE_T_463 (IN)	F24	mir-1277 microRNA precursor family (RF04154); mir-3688 microRNA precursor family (RF04161); mir-466 microRNA precursor family (RF04052); MIR862 microRNA precursor family (RF04187)
MITE_T_464 (IN)	F24	mir-1277 microRNA precursor family (RF04154); mir-3688 microRNA precursor family (RF04161); mir-466 microRNA precursor family (RF04052); MIR862 microRNA precursor family (RF04187)
MITE_T_2197 (DW)	F81	microRNA MIR828 (RF01026)
MITE_T_5650 (IG)	F1650	MIR7696 microRNA precursor family (RF03743)
MITE_T_8440 (DW)	F360	mir-8830 microRNA precursor family (RF03949); MIR1883 microRNA precursor family (RF03599)
MITE_T_8565 (IG)	F373	mir-297 microRNA precursor family (RF04217); mir-466 microRNA precursor family (RF04052)

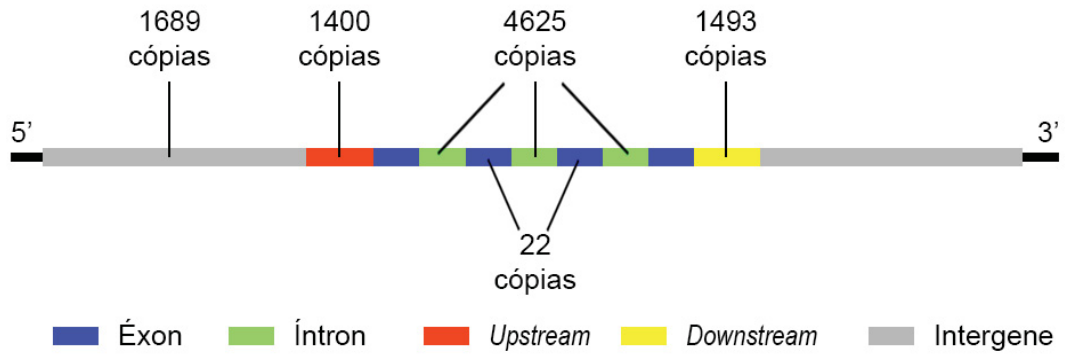
1 **Tabela 3** – miRNAs identificados em seqüências de MITES e sua localização no genoma de *Apareiodon* sp.

MITE (localização genômica)	Família	miRNA	miRNA idêntico em outras espécies	Gene (código gene; posição)
MITE_T_489 (IN)	F30	ola-miR-133-5p	dre-miR-133a-2-5p; dre-miR-133a-5p; ccr-miR-133a-5p; ssa-miR-133a-5p; gmo-miR-133a-5p	Splicing factor SWAP (g30936; intron 4)
MITE_T_494 (IG)	F31	ssa-miR-103-5p	-	-
MITE_T_553 (IN)	F33	dre-miR-203a-3p	fru-miR-203; tni-miR-203; ola-miR-203; pol-miR-203-3p; ccr-miR-203a; ssa-miR-203a-3p; abu-miR-203; mze-miR-203; nbr-miR-203; oni-miR-203a; pny-miR-203; gmo-miR-203a-3p	Collagen alpha-1 (XIX) (g21490; intron 5)
MITE_T_645 (IN)	F43	dre-let-7c-3p	pol-let-7d-3p; ssa-let-7c-3p; gmo-let-7c-3p	Sec1 family domain containing 2 (g17644; intron 3)
MITE_T_810 (DW)	F44	dre-miR-499-5p	ccr-miR-499; ipu-miR-499; ssa-miR-499b-5p; abu-miR-499; mze-miR-499; nbr-miR-499; oni-miR-499; pny-miR-499; gmo-miR-499-5p	Solute carrier organic anion transporter family, member 4A1 (g32558)
MITE_T_849 (IN)	F45	ssa-miR-462b-3p	-	Mbt domain containing 1 (g15115; intron 27)
MITE_T_958 (IN)	F51	oni-miR-10959	-	Membrane-spanning 4-domains subfamily A 4D (g6320; intron 1)
MITE_T_1048 (IN)	F51	oni-miR-10959	-	Ubiquitin protein ligase E3 component n-recognin (g6335; intron 55)
MITE_T_1471 (IN)	F65	gmo-miR-29a-3-5p	-	Kalirin RhoGEF kinase b (g30293; intron 1)
MITE_T_1634 (DW)	F71	oni-miR-10665	-	NADH:ubiquinone oxidoreductase subunit A11 (g22404)
MITE_T_1636 (IG)	F71	oni-miR-10665	-	-
MITE_T_1640 (IN)	F71	oni-miR-10665	-	Transmembrane protein FAM155A (g12596; intron 1)
MITE_T_2061 (DW)	F78	dre-miR-203a-3p	fru-miR-203; tni-miR-203; ola-miR-203; pol-miR-203-3p; ccr-miR-203a; ssa-miR-203a-3p; abu-miR-203; mze-miR-203; nbr-miR-203; oni-miR-203a; pny-miR-203; gmo-miR-203a-3p	Transient receptor potential cation channel subfamily M member 4-like (g15143)
MITE_T_2129 (IN)	F78	dre-miR-203a-3p	fru-miR-203; tni-miR-203; ola-miR-203; pol-miR-203-3p; ccr-miR-203a; ssa-miR-203a-3p; abu-miR-203; mze-miR-203; nbr-miR-203; oni-miR-203a; pny-miR-203; gmo-miR-203a-3p	Complement C3-like (g28346; intron 8)
MITE_T_2434 (IN)	F86	oni-miR-10637	-	Ovochymase-2 (g29897; intron 8)
MITE_T_2614 (IG)	F87	gmo-miR-29a-3-5p	-	-

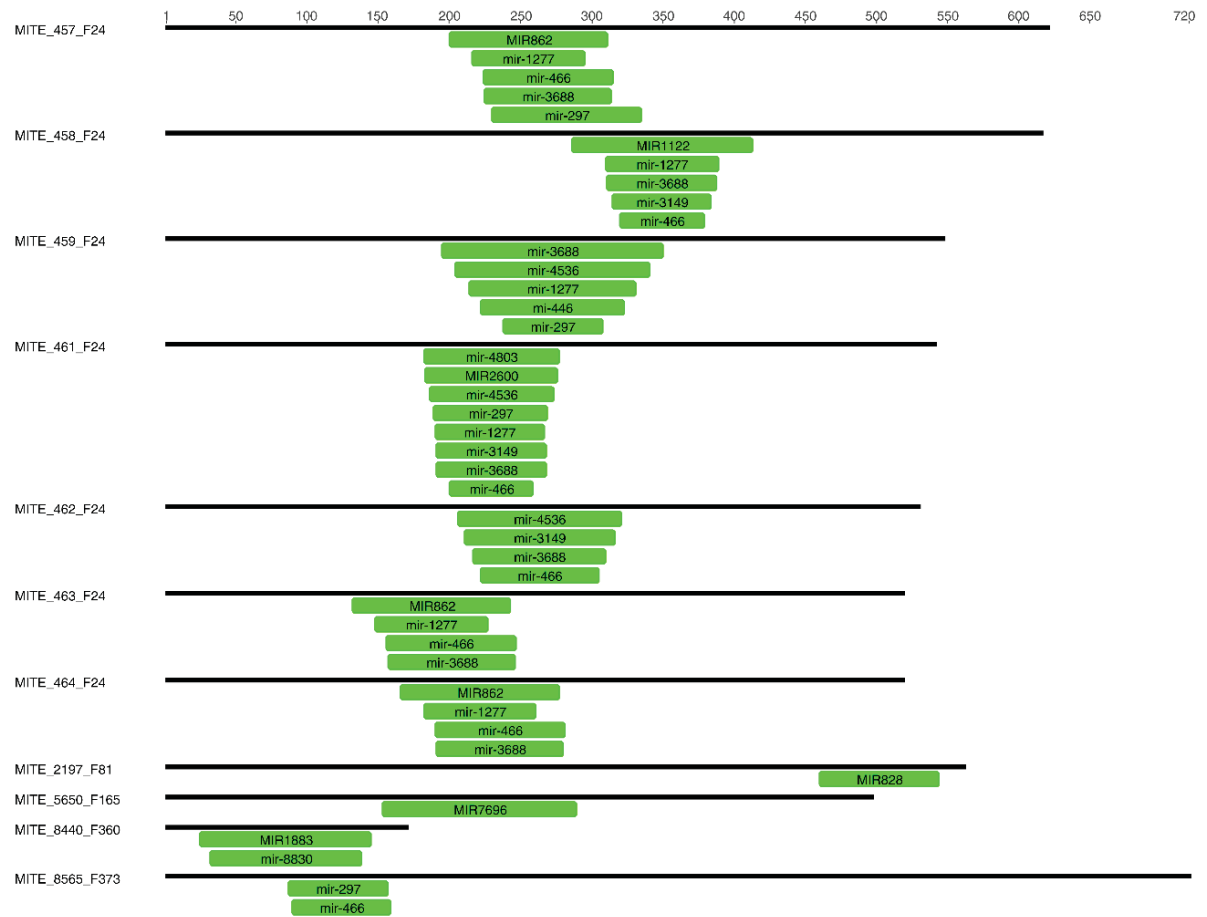
MITE (localização genômica)	Família	miRNA	miRNA idêntico em outras espécies	Gene (código gene; posição)
MITE_T_2663 (IN)	F87	gmo-miR-29a-3-5p	-	Exocyst complex component 1 (g17666; intron 12)
MITE_T_2811 (IN)	F87	gmo-miR-29a-3-5p	-	Histone acetyltransferase KAT5-like (g24538; intron 14)
MITE_T_2897 (DW)	F87	gmo-miR-9-5-3p	-	Adaptor related protein complex 2 subunit alpha 2 (g22948)
MITE_T_2921 (IN)	F88	oni-miR-10665	-	Glutamate receptor, ionotropic, N-methyl-D-aspartate 3Bb (g22369; intron 5)
MITE_T_3098 (IG)	F96	dre-miR-203a-3p	fru-miR-203; tni-miR-203; ola-miR-203; pol-miR-203-3p; ccr-miR-203a; ssa-miR-203a-3p; abu-miR-203; mze-miR-203; nbr-miR-203; oni-miR-203a; pny-miR-203; gmo-miR-203a-3p; ipu-miR-30a	-
MITE_T_3635 (IN)	F115	dre-miR-30a-3p		Chromodomain helicase DNA binding protein 3 (g9853; intron 23)
MITE_T_3649 (IN)	F117	oni-miR-10712	-	PHD finger protein 14 (g35317; intron 4)
MITE_T_3677 (IG)	F119	oni-miR-10838	-	-
MITE_T_4245 (IN)	F122	oni-miR-10577	-	Ceramide kinase-like (g11845; intron 5)
MITE_T_4832 (IN)	F137	oni-miR-10712	-	Insulin-like growth factor 1a receptor (g23214; intron 1)
MITE_T_5692 (IN)	F168	oni-miR-10675	-	v-akt murine thymoma viral oncogene homolog 3b (g28117; intron 4)
MITE_T_5767 (DW)	F174	oni-miR-10712	-	RAB38b (RAS oncogene) (g7745)
MITE_T_5791 (IG)	F174	oni-miR-10660	-	-
MITE_T_5801 (IG)	F177	gmo-miR-736-5p	-	-
MITE_T_6267 (IN)	F203	oni-miR-10864	-	Family with similarity 78 member Aa (g16516; intron 4)
MITE_T_6301 (DW)	F204	dre-miR-732	-	Vestigial-like family member 4b (g10829)
MITE_T_6471 (IN)	F231	oni-miR-10647	-	Proline-rich receptor-like protein kinase (g869; intron 11)
MITE_T_6749 (IN)	F270	oni-miR-726b	-	Transmembrane protein 151B (g25399; intron 1)
MITE_T_7079 (IG)	F318	dre-miR-3906	-	-
MITE_T_7136 (IN)	F324	oni-miR-10974	-	Coiled-coil domain containing 157 (g18533; intron 2)

MITE (localização genômica)	Família	miRNA	miRNA idêntico em outras espécies	Gene (código gene; posição)
MITE_T_7137 (IN)	F324	oni-miR-10974	-	Ectonucleotídeo pyrophosphatase/phosphodiesterase family member 2 (g19612; intron 18)
MITE_T_7138 (IN)	F324	oni-miR-10974	-	Amyloid beta (A4) precursor protein-binding (g1642; intron 21)
MITE_T_7139 (IN)	F324	oni-miR-10974	-	Tudor domain containing 9 (g3320; intron 2)
MITE_T_8338 (IN)	F350	oni-miR-10948	-	Thymocyte selection-associated high mobile group box (g5146; intron 6)
MITE_T_8765 (IN)	F389	oni-miR-10910	-	LARGE xylosyl- and glucuronyltransferase 1 (g702; intron 11)
MITE_T_8801 (IN)	F389	oni-miR-10910	-	LARGE xylosyl- and glucuronyltransferase 1 (g702; intron 16)
MITE_T_8912 (DW)	F398	ssa-miR-103-5p	-	Calmodulin-binding transcription activator 1 (g32857)
MITE_T_8914 (UP)	F398	ssa-miR-103-5p	-	Skin secretory protein xP2-like (g28552)
MITE_T_8915 (IG)	F398	ssa-miR-103-5p	-	-
MITE_T_9098 (DW)	F410	ssa-miR-103-5p	-	Solute carrier family 25 member 10 (g28337)
MITE_T_9101 (IN)	F410	ssa-miR-103-5p	-	G-protein coupled receptor 158 (g22340; intron 4)
MITE_T_9143 (IN)	F416	gmo-miR-93-3p	-	Cell adhesion molecule 4 (g34238; intron 4)
MITE_T_9188 (IG)	F420	dre-miR-153c-5p	-	-
MITE_T_9199 (IN)	F423	ipu-miR-7553	-	ArfGAP with GTPase domain, ankyrin repeat and PH domain (g34142; intron 15)

1 **Figura 1** – Identificação do local das MITEs no genoma de *Apareiodon* sp.



1 **Figura 2** – Identificação da similaridade das famílias de ncRNAs para as MITEs de *Apareiodon* sp. Nove
 2 MITEs apresentaram similaridade para duas ou mais famílias ncRNA.



3

1 4.2 CAPÍTULO 2

2

3

4 Major and minor U small nuclear RNAs genes characterization in a neotropical fish
5 genome: Chromosomal remodeling and repeats units dispersion in Parodontidae

6

7

8 Artigo publicado no jornal Gene (10.1016/j.gene.2022.146459)

1 **Major and minor U small nuclear RNAs genes characterization in a neotropical fish**
2 **genome: chromosomal remodeling and repeat units dispersion in Parodontidae**

3

4 Matheus Azambuja¹, Michelle Orane Schemberger¹, Viviane Nogaroto², Orlando Moreira-
5 Filho³, Cesar Martins⁴, Marcelo Ricardo Vicari^{1, 2}

6

7 ¹ Programa de Pós-Graduação em Genética, Universidade Federal do Paraná, Centro
8 Politécnico, Avenida Coronel Francisco H. dos Santos, 100, 81531-990, Curitiba, Paraná,
9 Brazil

10 ² Departamento de Biologia Estrutural, Molecular e Genética, Universidade Estadual de
11 Ponta Grossa, Av. Carlos Cavalcanti, 4748, 84030-900, Ponta Grossa, Paraná, Brazil

12 ³ Departamento de Genética e Evolução, Universidade Federal de São Carlos, Rodovia
13 Washington Luís, Km 235, 13565-905, São Carlos, São Paulo, Brazil.

14 ⁴ Departamento de Morfologia, Instituto de Biociências de Botucatu, Universidade
15 Estadual Paulista, Distrito de Rubião Júnior, s/n, 18618-689, Botucatu, São Paulo, Brazil.

16

17 **Abstract**

18 In association with many proteins, small nuclear RNAs (snRNAs) organize the
19 spliceosomes that play a significant role in processing precursor mRNAs during gene
20 expression. According to snRNAs genic arrangements, two kinds of spliceosomes (major
21 and minor) can be organized into eukaryotic cells. Although in situ localization of U1 and
22 U2 snDNAs have been performed in fish karyotypes, studies with genomic
23 characterization and functionality of U snRNAs integrated into chromosomal changes on
24 Teleostei are still scarce. This study aimed to achieve a genomic characterization of the
25 U snRNAs genes in *Apareiodon* sp. ($2n=54$ ZZ/ZW), apply these data to recognize
26 functional/defective copies, and map chromosomal changes involving snDNAs in
27 Parodontidae species karyotype diversification. Nine snRNA multigene families (*U1*, *U2*,
28 *U4*, *U5*, *U6*, *U11*, *U12*, *U4atac* and *U6atac*) arranged in putatively functional copies in the
29 genome were analyzed. Proximal Sequence Elements (PSE) and TATA-box promoters
30 occurrence, besides an entire transcribed region and conserved secondary structures,
31 qualify them for spliceosome activity. In addition, several defective copies or pseudogenes

1 were identified for the snRNAs that make up the major spliceosome. In situ localization of
2 snDNAs in five species of Parodontidae demonstrated that U1, U2, and U4 snDNAs were
3 involved in chromosomal location changes or units dispersion. The U snRNAs
4 defective/pseudogenes units dispersion could be favored by the probable occurrence of
5 active retrotransposition enzymes in the *Apareiodon* genome. The U2 and U4 snDNAs
6 sites were involved in independent events in the differentiation of sex chromosomes
7 among Parodontidae lineages. The study characterized U snRNA genes that compose
8 major and minor spliceosome in the *Apareiodon* sp. genome and proposes that their
9 defective copies trigger chromosome differentiation and diversification events in
10 Parodontidae.

11

12 **Keywords:** snRNAs, secondary structure, FISH, sex chromosomes, fish genome
13 evolution.

14

15 **Introduction**

16

17 Eukaryotic chromosomes are composed of different classes of DNA, which can
18 be arranged only once in the genome (single copy DNA), or sequences that repeat
19 thousands or millions of times (repetitive DNA) (Sumner, 2003; López-Flores and Garrido-
20 Ramos, 2012). The repetitive DNA sequences are a large part of all eukaryotic genomes,
21 representing more than half of the total content of the DNA in most species (Biscotti et al.,
22 2015). In numerous organisms, repetitive sequences have been contributed to a sizable
23 portion of the karyotype variations observed (Kidwell, 2002; Biscotti et al., 2015).

24 The multigene families are a group of genes with similar functions and sequences
25 that originate from a common ancestral gene (Nei and Rooney, 2005). Gene duplication
26 represents the core process in the evolution of multigene families since enormous
27 quantities of their products are required in cells (Sumner 2003). Nonetheless, other
28 phenomena, such as pseudogenization, loss of genes, recombination, and natural
29 selection act to varying degrees to shape the evolution of multigene families (Eirín-Lopez
30 et al., 2012).

1 Multigene families can show variable distribution patterns in karyotypes of
2 different species (Gornung, 2013; Anjos et al., 2015; Vitales et al., 2017; Sochorová et al.,
3 2018; Degrandi et al., 2020). In related species, they can be located on homeologous
4 chromosomes or present low rates of chromosome changes (Vicari et al., 2006a;
5 Milhomem et al., 2013; Souza e Sousa et al., 2017; Ponzio et al., 2018; Gazoni et al.,
6 2021). Conversely, chromosomal sites surrounding multigene families have been
7 described as prone to double-strand breaks and to generate highly reorganized
8 karyotypes (Kehrer-Sawatzki and Cooper, 2008; Cazaux et al., 2011; Maneechot et al.,
9 2016; Cavalcante et al., 2018; Glugoski et al., 2018; Deon et al., 2020; Takagui et al.,
10 2020).

11 Large amounts of transcripts in eukaryotic cells are synthesized in distinct types
12 of non-coding RNAs (ncRNAs) (Amaral et al., 2008; Liu et al., 2013; Deveson et al., 2017).
13 According to their functions, the ncRNAs can be assembled into two groups:
14 housekeeping ncRNAs, which are abundant in cells, and primarily regulate generic
15 cellular processes; and regulatory ncRNAs, which function as regulators of gene
16 expression at epigenetic, transcriptional, and post-transcriptional levels (Ponjavic et al.,
17 2007; Cech and Steitz, 2014; Peschansky and Wahlestedt, 2014; Zhang et al., 2019).

18 A massive portion of genes transcribed by RNA polymerase II is expressed as
19 precursor mRNAs (pre-mRNAs). Posteriorly, they are processed into mature mRNAs by
20 splicing, where noncoding sequences (introns) are removed, and coding sequences
21 (exons) are connected, an essential step of gene expression in eukaryotic tissues (Will
22 and Lührmann, 2011). The spliceosome, complex machinery responsible for catalytic
23 splicing, involves five different subunits of ribonucleoproteins (RNP) associated with
24 protein cofactors (Lerner et al., 1980; Will and Lührmann, 2011; Shi, 2017). Two kinds of
25 spliceosomes could be found in eukaryotes: the U2-dependent spliceosome (major
26 spliceosome), which catalyzes U2-type introns, and a less abundant U12-dependent
27 spliceosome (minor spliceosome) which removes the rare U12-type introns (Patel and
28 Steitz, 2003; Turunen et al., 2013). Multiple proteins and *U1*, *U2*, *U4*, *U5*, and *U6 snRNAs*
29 organize the major spliceosome, while the minor spliceosome assembling uses the *U5*
30 *snRNA* (a common snRNA between major and minor spliceosome) and the *U11*, *U12*,
31 *U4atac*, and *U6atac snRNAs*, which have a similar structure and function to the *U1*, *U2*,

1 *U4*, and *U6 snRNAs* (Patel and Steitz, 2003). U12-type introns generally represent less
2 than 0.5% of all introns in the genomes (Turunen et al., 2013). They differ from U2-type
3 introns by the consensus sequence that is recognized by the spliceosome: while U2-type
4 introns have GT-AG termini (rarely AT-AC and GC-AG), U12-type has AT-AC termini
5 (rarely GT-AG) (Patel and Steitz, 2003; Sheth et al., 2006).

6 From a cytogenetic point of view, contrary to former studies for rDNA clusters
7 widely mapped on fish chromosomes (Gornung, 2013; Sochorová et al., 2018), snRNAs
8 are poorly explored, and their clusters characterizations at genome level are lacking. Still,
9 the U1 and U2 snDNAs were helpful to understand the evolutionary relationships among
10 closely related species (Merlo et al., 2010, 2013; Úbeda-Manzanaro et al., 2010; Cabral-
11 de-Mello et al., 2012; Utsunomia et al., 2014; Scacchetti et al., 2015; Silva et al., 2015;
12 Piscor et al., 2016, 2018; Yano et al., 2017, 2020; Ponzio et al., 2018; Pucci et al., 2018;
13 Sember et al., 2018; Dulz et al., 2020; Malimpensa et al., 2020).

14 Parodontidae (Teleostei) gathers species with a diploid number (2n) of 54
15 chromosomes, usually bi-armed, but with the presence of acrocentrics in some species
16 (Bellafronte et al., 2011, 2012; Traldi et al., 2016, 2020; Santos et al., 2019; Nirchio et al.,
17 2021). Despite the shared 2n, some species are distinguished according to chromosomal
18 characteristics: karyotypic formula, accumulation of heterochromatic regions, presence or
19 absence of heteromorphic sex chromosomes, number of sites of 18S and 5S rDNAs,
20 distribution of the DNA satellite pPh2004, *WAp* repetitive portion, and transposable
21 elements (Bellafronte et al., 2011, 2012; Schemberger et al., 2011, 2014, 2016, 2019;
22 Ziemniczack et al., 2014; Traldi et al., 2016, 2020; Nascimento et al., 2018; Santos et al.,
23 2019; Nirchio et al., 2021). In Parodontidae, the remarkable chromosomal feature is the
24 differentiation of heteromorphic sex chromosomes in some species (Schemberger et al.,
25 2011, 2019). Their species may lack heteromorphic sex chromosomes, show a
26 recognizable proto-sex chromosome pair, a ZZ/ZW sex chromosome system, or possess
27 a ZZ/ZW₁W₂ multiple sex chromosome system, allowing chromosomal evaluations of
28 differentiation and diversification in the group (Moreira-Filho et al., 1980, 1993; Centofante
29 et al., 2002; Rosa et al., 2006; Vicari et al., 2006b; Bellafronte et al., 2011, 2012;
30 Schemberger et al., 2011; Traldi et al., 2016, 2020; Nascimento et al., 2018; Santos et al.,
31 2019; Nirchio et al., 2021).

1 *Apareiodon* sp. (an undescribed species in scientific literature and closely related
2 to *A. ibitiensis*) possesses a $2n=54$ and a ZZ/ZW sex chromosome system (Vicari et al.,
3 2006b). Recently, this species had its genome assembled (Schemberger et al., 2019),
4 allowing integrated genomic and chromosomal data analysis. This study aimed to perform
5 the prediction and characterization of U snRNAs in the *Apareiodon* sp. genome seeking
6 to understand mechanisms of evolution and dispersion of their sequences. Furthermore,
7 the snRNAs obtained were used to demonstrate the role of these multigene families in
8 chromosomal rearrangements, especially those involved in the
9 differentiation/diversification of sex chromosomes in Parodontidae species.

11 **Material and methods**

13 ***Prediction and characterization of snRNAs in Apareiodon sp.***

15 A prediction of snRNAs belonging to the major and minor spliceosome was made
16 using Infernal 1.1.3 (Nawrocki and Eddy, 2013), employing the Rfam (v. 14.2 – April 2020)
17 (Kalvari et al., 2020) library of covariance models (U1 – RF00003; U2 – RF00004; U4 –
18 RF00015; U5 – RF00020; U6 – RF00026; U11 – RF00548; U12 – RF00007; U4atac –
19 RF00618; U6atac – RF00619) according to Alternate Protocol 1 proposed by Kalvari et
20 al. (2018) in the female genome assembly of *Apareiodon* sp. (Schemberger et al., 2019)
21 (available at [http://sacibase.ibb.unesp.br/jbrowse/JBrowse-](http://sacibase.ibb.unesp.br/jbrowse/JBrowse-1.12.1/index.html?data=dataapareiodon-sp)
22 [1.12.1/index.html?data=dataapareiodon-sp](http://sacibase.ibb.unesp.br/jbrowse/JBrowse-1.12.1/index.html?data=dataapareiodon-sp)).

23 Secondary structures of the *U1*, *U2*, *U4*, *U5*, and *U6 snRNAs* were obtained after
24 aligning the ten sequences with lower E-values using CLUSTALW in Geneious v 7.1.9
25 software (Kearse et al., 2012). All predicted sequences of the *U11*, *U12*, *U4atac* and
26 *U6atac snRNAs* were aligned to compute their secondary structures. The RNAalifold tool
27 (Bernhart et al., 2008) available in the ViennaRNA package 2.0 (Lorenz et al., 2011) was
28 used to obtain the secondary structures. The heterodimer formation between the *U4/U6*
29 and *U4atac/U6atac snRNAs* was generated from their consensus sequences, and the
30 RNAcofold tool was used to determine the interaction (Bernhart et al., 2006).

31

1 ***Identification and characterization of promoters***

2

3 The MEME suite software v. 5.3.3 (Bailey et al., 2009), which includes the MEME
4 tool (Number of motifs per sequence = 0 or 1; the maximum number of motifs = 10; motifs
5 wide = 5–25; the number of sites in each motif = 5–20), was applied to discover conserved
6 motifs in segments up to 200 bp upstream in the prospected snRNA sequences. After
7 identifying the motifs, the search for the conserved Proximal Sequence Element (PSE)
8 and TATA-box elements was performed manually in the snRNAs sequences.

9

10 ***Primer design and validation of snRNAs in the *Apareiodon* sp. genome***

11

12 Primers for each snRNA family were designed by aligning predicted sequences
13 of the *Apareiodon* sp. genome with lower E-values in Primer3Plus software (Untergasser
14 et al., 2007) as described in Table S1 - Electronic supplementary material. The genomic
15 DNA of *Apareiodon* sp. was extracted from liver samples following the CTAB method
16 (Murray and Thompson, 1980) and used as a template for Polymerase Chain Reactions
17 (PCR) to obtain each snRNA amplicon. Reaction mix contained: 1 x reaction buffer (200
18 mM Tris pH 8.4, 500 mM KCl), 1 mM MgCl₂, 0.2 mM dNTPs, 0.5 μM of each primer, 1 U
19 *Taq* DNA polymerase (5 U/μL), and 40 ng of DNA template. The reaction program used
20 consists of initial denaturation for 10 min at 95°C, 35 cycles of 95°C for 1 min, 61,5°C for
21 45 s and 72°C for 1 min, and a final extension at 72°C for 10 min. The PCR products were
22 purified with the Illustra GFX PCR DNA and Gel Band Purification (GE Healthcare,
23 Chicago, USA), then cloned with the pGEM®-T Easy Vector Systems Kit (Promega,
24 Madison, USA) in *Escherichia coli* DH5α and sequenced in the platform ABI-prism 3500
25 Genetic Analyzer (Applied Biosystems). The sequences were analyzed in the software
26 Geneious v 7.1.9 (Kearse et al., 2012), and identities were determined using the tools
27 Nucleotide Basic Local Alignment Search Tool (BLASTn) (Altschul et al., 1990), Rfam
28 (Kalvari et al., 2020), and CENSOR (Kohany et al., 2006).

1 **Chromosome preparation and fluorescent in situ hybridization (FISH)**

2

3 Mitotic chromosomes of five species of Parodontidae (Table S2 - Electronic
4 supplementary material) were obtained from the anterior kidney by the air-drying method
5 (Bertollo et al., 2015). The *U1*, *U2*, *U4*, *U5*, and *U6 snRNAs* sequences were amplified as
6 probes via PCR using digoxigenin-11dUTP (Jena Bioscience, Jena, Germany) (*U1*),
7 Aminoallyl-dUTP-Cy5 (Jena Bioscience, Jena, Germany) (*U2* and *U4*), and biotin-16-
8 dUTP (Jena Bioscience, Jena, Germany) (*U5* and *U6*), and used in the in situ localization
9 experiments. Fluorescence in situ hybridization (FISH) was performed following the
10 general protocol described by Pinkel et al. (1986), and the hybridization step was
11 conducted under stringency conditions as follow: 200 ng of each probe, 50% formamide,
12 10% dextran sulfate, and 2xSSC – saline-sodium citrate; 16 h of hybridization at 37°C.
13 Signals detection was performed using Streptavidin Alexa Fluor 488 (Molecular Probes,
14 Eugene, USA) and anti-digoxigenin rhodamine Fab fragments (Roche Applied Science,
15 Penzberg, Germany) for the *U1*, *U5* and *U6 snDNAs* probes. Chromosomes were counter-
16 stained with 0.2 µg/mL of DAPI in VECTASHIELD mounting medium (Vector Laboratories,
17 Burlingame, USA). Metaphases submitted to in situ localization were analyzed in an
18 epifluorescence microscope (Leica DM 2000) coupled to a DFC3000 G CCD camera
19 (Leica). Chromosomes were classified as metacentric (m), submetacentric (sm), and
20 subtelocentric (st) according to the arm ratio rule (Levan et al., 1964), and arranged into
21 karyotypes.

22

23 **Results**

24

25 **Prediction and genomic characterization of snRNAs**

26

27 A total of 536 sequences were identified into nine different U snRNAs in the
28 *Apareiodon* sp. genome, which were grouped as follow: 295 *U1 snRNA*, 72 *U2 snRNA*,
29 15 *U4 snRNA*, 29 *U5 snRNA*, 113 *U6 snRNA*, 4 *U11 snRNA*, 2 *U12 snRNA*, 2 *U4atac*
30 *snRNA*, and 4 *U6atac snRNA*. The *U1 snRNA* presented 164 nucleotides (nt) in length,
31 and the structure obtained demonstrated the formation of four stem-loops (Figure 1a). A

1 sequence with 191 nt was observed in the *U2 snRNA*, and the predicted structure also
 2 presented four stem-loops (Figure 1b). The *U4 snRNA* had 141 nt and showed a long
 3 stem-loop besides some minor loops in the structure (Figure 1c). The *U5 snRNA* had 116
 4 nt, and its secondary structure was arranged with two stem-loops and one stem (Figure
 5 1d). The *U6 snRNA* presented a sequence with 107 nt, and its structure demonstrated the
 6 occurrence of a stem-loop, a stem, and some loops (Figure 1e). The components of the
 7 minor spliceosome were characterized as follows: the *U11* and *U12 snRNAs* showed 135
 8 nt and 153 nt, respectively, and a secondary structure with four stem-loops (Figure 1f, g);
 9 the *U4atac snRNA* was 132 nt long and a stem-loop and some loops were observed in its
 10 structure (Figure 1h); at last, a molecule 129 nt long organized in stem-loops and one
 11 stem was observed in the *U6atac snRNA* (Figure 1i). Secondary structures arranged with
 12 stem-loops and some loops were observed in the U4/U6 and U4atac/U6atac complexed
 13 snRNAs (Figure 2).

14 Conserved motifs containing promoter sequences were observed in all predicted
 15 snRNA genes in the *Apareiodon* sp. genome. The PSE was observed in all snRNAs, while
 16 the TATA-box consensus was visualized only in the *U6* and *U6atac snRNAs* (Figure 3).
 17 The *U1 snRNA* PSE consensus 5'-CAACCCMNCVACTGCNCCSKV-3' was observed in
 18 20 sequences ranging between genic positions -62 and -42 (Figure 3a). The PSE
 19 consensus for *U2 snRNA* was 19 nt long (5'-TCACCTAGTGGGGAGCTCT-3') and was
 20 located at the genic region -65 to -46 in nine sequences (Figure 3b). A PSE consensus
 21 (5'-RGTAACCGTTCCTGTGATGTC-3') between the -63 to -43 genic region was
 22 observed in eight *U4 snRNA* sequences (Figure 3c). Seven *U5 snRNA* presented a PSE
 23 consensus (5'-TAACCGTCCAGTAGGAATTTAAGT-3') 24 nt long located between at the
 24 -66 and -43 genic region (Figure 3d). The *U6 snRNA* PSE consensus (5'-
 25 TCTACCACGGACAAGAACAATA-3') was observed at the region -66 and -46 upstream
 26 of the transcribed region (Figure 3e). In *U6 snRNA*, the TATA-box consensus (5'-
 27 TACTTAAA-3') was located between positions -31 and -24 upstream of the transcribed
 28 region and determined in 12 sequences of the *U6 snRNA* (Figure 3j).

29 In minor spliceosome snRNAs genes, the PSE was also identified. The *U11*
 30 *snRNA* PSE (5'-GTTTCCCCTCCCGGAGAAGAGC-3') was located in genic regions -61
 31 and -40 (Figure 3f). The PSE consensus 5'-CYYACCGTCKCTGTAGTTCG-3' was

1 detected in two *U12 snRNA* at genic regions -62 to -43 (Figure 3g). The PSE consensus
 2 (5'-KTMACCRGCCWGTAGYYTMC-3') was located in the genic region -62 and -43 for
 3 two *U4atac snRNA* sequences (Figure 3h). *U6atac snRNA* has a 24 nt long PSE
 4 sequence (5'-GCCACCGTCTTTGATACACGTCTG-3') starting at position -67 (Figure 3i)
 5 and a TATA-box (5'-TTTATATT-3') at the position -31 to -24 (Figure 3k).

6 Sanger sequencing of the U snRNAs genes was obtained and compared with
 7 those prospected in the *Apareiodon* sp. genome. When submitted to the GenBank
 8 database, all sequences showed similarity with fish snRNAs and were confirmed as
 9 snRNAs in the Rfam database (Table S3 - Electronic supplementary material). *U6 snRNA*
 10 sequence showed an inner 52 bp-long segment 86% similar to DNA Transposon (DNA-
 11 8N1-Hsal) (Table S3 - Electronic supplementary material). Sequences were deposited as
 12 vouchers in GenBank (MZ645211 – MZ645219).

13 14 ***In situ localization***

15
16 The in situ localization of U1, U2, U4, U5, and U6 snDNAs was performed in five
 17 Parodontidae species (Figure 4). In all species analyzed, the U1 snDNA was located at
 18 the terminal region of m pair 1, corresponding to the Z chromosome in *A. affinis* (Figures
 19 4a, b, c, d, e). In addition, two more chromosome pairs presented U1 snDNA sites, as
 20 follows: in *A. affinis*, at the q proximal and the centromeric regions in the sm pairs 19 and
 21 20, respectively (Figure 4a); *Apareiodon piracicabae* had U1 snDNA sites located at the
 22 centromeric and at the q proximal regions in the sm pairs 20 and 21, respectively (Figure
 23 4b); *Apareiodon* sp. presented sites at the q proximal regions in the sm pairs 19 and 20
 24 (Figure 4c); in *P. hilarii* the centromeric region of sm pair 5 and the q proximal region of
 25 sm pair 22 showed signals (Figure 4d); and, *P. nasus* demonstrated sites for U1 snDNA
 26 at the q proximal and centromeric regions of sm pairs 4 and 7, respectively (Figure 4e).

27 The U2 snDNA was mapped at the centromeric region of sm pair 2 and the q
 28 distal region of st pair 27 in *A. affinis* (Figure 4f). *Apareiodon piracicabae* presented *U2*
 29 *snDNA* sites at the proximal region of the sm pair 5 and the q interstitial region of the sm
 30 pair 7 (Figure 4g). The U2 snDNA sites were located at the centromeric region of the m
 31 pair 2 and on the sex chromosomes Z and W in *Apareiodon* sp. (Figure 4h). In *P. hilarii*,

1 the *U2 snDNA* was at the q distal region on sm pair 4 and the centromeric region of the
2 m pair 8 (Figure 4i). In *P. nasus*, it was located at the centromeric region of the sm pair 8
3 (Figure 4j).

4 For the *U4 snDNA*, all analyzed species showed signals at the centromeric region
5 of m pair 1 (Figures 4k, l, m, n, o), corresponding to the Z chromosome in *A. affinis* (Figure
6 4k). Besides that, *U4 snDNA* sites were also observed at the terminal region of the Z
7 chromosome and centromeric region of the W_1W_2 in *A. affinis* (Figure 4k), at the terminal
8 region of pair 1, and the q distal region of the st pair 27 in *A. piracicabae* (Figure 4l), at
9 the terminal region of the pair 1 in *Apareiodon* sp. (Figure 4m) and accumulated along the
10 W q arm in *P. hilarii* (Figure 4n).

11 The *U5 snDNA* was detected in only one chromosome pair in all analyzed
12 species, as follows: q distal region of the chromosome pair 5 in *A. affinis* (Figure 4p), and
13 at the q proximal region of the chromosome pairs 8, 3, 7, and 5 in *A. piracicabae*,
14 *Apareiodon* sp., *P. hilarii*, and *P. nasus*, respectively (Figure 4p, q, r, s, t).

15 The *U6 snDNA* showed a single site in all species analyzed, located at the
16 proximal region of the chromosome pairs 21, 16 and 19 in *A. affinis*, *Apareiodon* sp. and
17 *P. nasus*, respectively (Figure 4u, w, y), on the q interstitial region of the sm pair 24 in *A.*
18 *piracicabae* (Figure 4v), and in the chromosome pair 15 in *P. hilarii* (Figure 4x).

19

20 Discussion

21

22 *snRNAs genes in Apareiodon sp. genome*

23

24 The major spliceosome is the dominant splicing machinery (Sheth et al., 2006).
25 The minor spliceosome is present in many lineages and has an early origin in eukaryotic
26 diversification (Collins and Penny, 2005; Lorković et al., 2005; Russel et al., 2006; López
27 et al., 2008). Components of the minor spliceosome could be not present in some groups
28 (e.g., yeast and nematodes), and in other cases, some minor spliceosome snRNA genes
29 could be absent (Mewes et al., 1997; Burge et al., 1998; Russel et al., 2006; López et al.,
30 2008; Marz et al., 2008). In the *Apareiodon* sp. genome, the nine snRNAs genes involved

1 in the organization of the major and minor spliceosomes were annotated by bioinformatics
2 tools and validated by Sanger sequencing and molecular structure characterization.

3 In fishes, all U snRNA genes were annotated in the genome of *Fugu rubripes*,
4 *Gasterosteus aculeatus*, *Oryzias latipes*, *Petromyzon marinus*, and *Tetraodon nigroviridis*
5 (López et al., 2008). In contrast, in *Danio rerio*, the *U12 snRNA* was not predicted (López
6 et al., 2008). However, when analyzing the same six species, Marz et al. (2008) identified
7 all functional snRNAs, except the *U4atac snRNA* in *T. nigroviridis*. Many ncRNAs may
8 have low nucleotide sequence conservation, making it challenging to recognize homology
9 based only on similarity (Kalvari et al., 2018). Thus, they are typically characterized by a
10 specific secondary structure (López et al., 2008; Kalvari et al., 2018). snRNAs sequences
11 tend to be conserved during evolution compared to other ncRNAs (López et al., 2008),
12 but a combination of methods must be employed to maximize prediction results (López et
13 al., 2008; Kalvari et al., 2018).

14 Secondary structures for ncRNAs in fish are rare and limited to rRNAs (Barman
15 et al., 2016; Barros et al., 2017; Cao et al., 2020), with few exceptions (Merlo et al., 2013;
16 Chairi and Gonzales, 2015). The secondary structures obtained for snRNAs in
17 *Apareiodon* sp. were similar to those proposed by Marz et al. (2008) and found for some
18 fish and several other vertebrate groups. This structure conservation, arranged in stem-
19 loops and loops, corroborates the proposition of functionalization of these snRNAs in
20 *Apareiodon* sp. cells. The structure complexed for *U4/U6 snRNAs* and their
21 *U4atac/U6atac* analogs reinforce this functionalization in *Apareiodon* sp. In *U4/U6.U5* tri-
22 snRNP, *U6 snRNA* is paired with the *U4 snRNA* as an essential step to intron removal,
23 acting as a chaperone to keep U6 in a precatalytic conformation, and it prevents pre-
24 pairing between *U2* and *U6 snRNAs* (for a review, see Wilkinson et al., 2020).

25 Once some snRNA pseudogenes present enough similarity to the functional
26 copies from which they have derived, a severe limitation of Infernal software is its inability
27 to distinguish pseudogenes from the functional ncRNA genes (Kalvari et al., 2018).
28 Besides that, defective copies derived from multigene families are familiar to genomes
29 (Ciganda and Williams, 2011). A quantity of *U1*, *U2*, and *U6 snRNAs* sequences was
30 recovered from the *Apareiodon* sp. genome, indicating the possible occurrence of

1 defective/pseudogene copies among them, which was demonstrated by the absence of
2 promoter sequences in many of these copies.

3 Pseudogenes can be generated by the degeneration of single-copy genes or
4 multigene units (Li et al., 2013). In another pathway, pseudogenes could have genomic
5 spread by retrotransposing their RNAs molecules (Li et al., 2013). Thus, they are
6 categorized into three major groups: unitary pseudogenes, duplicated pseudogenes, and
7 processed pseudogenes (for a review, see Li et al., 2013). The absence of a 5'-promoter
8 sequence typically diagnoses processed pseudogenes or retro-pseudogenes, sometimes
9 possessing a 3'-end poly(A) sequence and direct repeats in the flanks (Mighell et al., 2000;
10 Li et al., 2013; Kovalenko and Patrushev, 2018). Different classes of snRNAs retro-
11 pseudogenes were identified in several mammalian genomes and suggest that
12 retrotransposition pathways may be multiple and specific to each genomic environment,
13 especially when it involves long interspersed nuclear elements (LINEs) (Doucet et al.,
14 2015). Furthermore, Kojima (2015) described a new class of short interspersed nuclear
15 elements (SINE), the SINEU, whose heads originate from *U1* and/or *U2 snRNAs* in
16 crocodylian genomes.

17 Sequence analysis of snRNAs and their flanking regions in fish demonstrated
18 their association with different transposable elements (Cabral-de-Mello et al., 2012; Pucci
19 et al., 2018; Dulz et al., 2020; Malimpensa et al., 2020; Yano et al., 2020), showing that
20 these elements may be involved in the dispersion of snRNA sequences. In addition,
21 Kojima and Jurka (2013) reported the *Dada* transposons, a superfamily of DNA
22 transposons whose members have been found inserted into specific types of small RNA
23 genes, such as the *U1* and *U6 snRNAs* genes in some fish genomes.

24 snRNAs have a unique array of promoter sequences, which include a Distal
25 Sequence Element (DSE), which acts as an enhancer, and a proximal sequence element
26 (PSE), and in some genes, a TATA-box consensus may be located adjacent to the PSE
27 (Hernandez, 2001; Jawdekar and Henry, 2008). The PSE and TATA-box elements tend
28 to be more conserved among the snRNAs sequences than other snRNA-associated
29 upstream elements (Marz et al., 2008). For example, SPH motif, octamer (OCT), CAAT-
30 box, GC-box, -35-element, and initiator (Inr), are even less conserved between different
31 genes in the same species and between homologous genes in distinct species (Marz et

1 al., 2008). The mutual presence of PSE and TATA elements directs to the recruitment of
2 the specific mechanism of RNA polymerase III, while the absence of TATA-box specifies
3 the recruitment of the transcriptional apparatus of RNA polymerase II (Hernandez, 2001;
4 Jawdekar and Henry, 2008). In *Apareiodon* sp., *U6* and *U6atac snRNAs* present the two
5 main elements for the recruitment of RNA polymerase III, while the other snRNAs are
6 transcribed by RNA polymerase II. All data of snRNAs characterized in the *Apareiodon*
7 sp. genome demonstrated functional genes for major and minor spliceosomes besides
8 the presence of pseudogenes/defective copies.

9 10 ***snRNAs and sex chromosomes differentiation***

11
12 Chromosomal markers could be represented by repetitive DNAs helpful in
13 studying the structure and function of chromosomes, evolution, identification of
14 chromosomal rearrangements, and sex chromosomes (Martins et al., 2010; Vicari et al.,
15 2010). The in situ locations of snDNAs are often restricted to U1 and U2 snDNAs, but few
16 studies have been carried out to map all these sequences in Neotropical fish (Paim et al.,
17 2018). The U4 and U6 snDNAs were mapped only in the fish *Hollandichthys multifasciatus*
18 (Soares et al., 2021), while the U5 snDNA has never been used in chromosomal mapping
19 studies.

20 Here, we designed new sets of primers to amplify fragments of the transcribed
21 region of the snRNAs genes in fishes and performed the in situ localization of the U1, U2,
22 U4, U5, and U6 snDNAs as specific probes. Defective copies of snRNAs could have parts
23 of transposable elements and satellite DNAs, generating unspecific sites in the karyotypes
24 (Dulz et al., 2020). The high similarity of the snRNA transcribed regions here obtained and
25 the delimitation just these segments to be used as probes allow a reliable location of their
26 chromosomal sites. In the sequences obtained to confirm the presence of snRNAs in the
27 genome of *Apareiodon* sp., only the *U6 snRNA* sequence showed similarity (52 bp) to
28 repetitive DNA. However, this segment may represent an artifact of sequence or
29 evolutionary convergence since the TE DNA-8N1_Hsal is 816 bp-long, possess an
30 internal element similar to *U6 snRNA* and was described in the ant *Harpegnathos saltator*
31 (Bao and Jurka, 2014).

1 The five species analyzed cytogenetically have different karyotypic organizations
2 regarding the presence of sex chromosome systems: $2n = 54$, ZZ/ZW in *Apareiodon* sp.
3 and *Parodon hilarii*; $2n = 54♂/55♀$, ZZ/ZW₁W₂ in *Apareiodon affinis*; $2n = 54$ and proto-
4 sex chromosomes (pair 13) in *P. nasus*; $2n = 54$ and no heteromorphic sex chromosome
5 or evidence of proto-sex chromosomes in *Apareiodon piracicabae* (Vicari et al., 2006b;
6 Bellafronte et al., 2011; Schemberger et al., 2011; Nascimento et al., 2018). The snDNA
7 sites in Parodontidae karyotypes show no syntenic disposition of distinct kinds of snRNAs
8 (i.e., independently located). Most snDNAs showed a single-locus autosomal localization,
9 but some snRNA families showed multiple clusters in some Parodontidae species, in
10 addition to snDNAs sites on the ZW chromosomes that may be associated with the
11 differentiation or diversification of sex chromosomes in the group.

12 Fish species generally have a single U1 snDNA locus in the karyotype (Cabral-
13 de-Mello et al., 2012; Ponzio et al., 2018; Malimpensa et al., 2020; Yano et al., 2020;
14 Soares et al., 2021). However, multiple locations or scattered signals were also observed
15 (Manchado et al., 2006; García-Souto et al., 2015; Silva et al., 2015; Dulz et al., 2020). In
16 Parodontidae karyotypes, the U1 snDNA showed a variation in sites number. Although
17 multiple U1 snDNA sites have been found, the karyotypic comparison suggests
18 chromosomal sites location conservatism in Parodontidae lineage.

19 A probable homeologous U2 snDNA site in the proximal region of a bi-armed
20 chromosome was mapped in Parodontidae. Moreover, an additional locus with the
21 variable chromosomal location was detected among the species *Apareiodon* sp., *A.*
22 *affinis*, *A. piracicabae*, and *P. hilarii*. The single-locus condition for the U2 snDNA is the
23 most widely reported in fish species (Merlo et al., 2010, 2012a, 2013; Supiwong et al.,
24 2013; Utsunomia et al., 2014; García-Souto et al., 2015; Araya-Jaime et al., 2017; Serrano
25 et al., 2017; Piscor et al., 2018; Ponzio et al., 2018; Pucci et al., 2018; Malimpensa et al.,
26 2018, 2020; Sember et al., 2018; Dulz et al., 2020; Oliveira et al., 2020; Soares et al.,
27 2021). Sometimes, the U2 snDNA can be located in multiple sites (Manchado et al., 2006;
28 Utsunomia et al., 2014; Silva et al., 2015; Piscor et al., 2016; Yano et al., 2020) or
29 scattered across chromosomes (Úbeda-Manzanaro et al., 2010; Merlo et al., 2012b; Ráb
30 et al., 2016; Hatanaka et al., 2018), demonstrating a high transposition dynamism.

1 The dispersion of U1 and U2 snDNA sites in the fish chromosomes occurred
2 several times during the groups' evolution. Related species can present different
3 distribution patterns of snDNA sites, some probably lineage-specific, as observed in
4 Parodontidae. In addition, copies of U1 and U2 snDNAs have already been reported in
5 the sex chromosomes of the Neotropical fishes' genera *Gymnotus* and *Triportheus*
6 (Utsunomia et al., 2014; Yano et al., 2017). The U2 snDNA sequences present on sex
7 chromosomes only in *Apareiodon* sp. here analyzed reflect a specific event of invasion
8 and differentiation in the ZW chromosomes in Parodontidae.

9 *Apareiodon* lineage keeps two U4 snDNA sites in chromosome pair 1, besides an
10 additional locus observed in *A. piracicabae*. In *A. affinis* (ZZ/ZW₁W₂), chromosome pair 1
11 corresponds to chromosome Z, and homologous U4 snDNA sites were found in the male
12 karyotype (ZZ). Moreover, the U4 snDNA sites were arranged in Z, W₁, and W₂
13 chromosomes in the female karyotype. A translocation between autosomes and proto-sex
14 chromosomes was proposed as the first step during the evolution of the multiple sex
15 chromosome system in *A. affinis* (Schemberger et al., 2011). In this pathway, the Z would
16 then undergo centric fission followed by pericentric inversions to form the W₁ and W₂
17 chromosomes, as has been previously proposed by Moreira-Filho et al. (1980). Based on
18 the homeology for U1 and U4 snDNA sites for chromosome 1 in *Apareiodon* lineage, it
19 was possible to suggest that U4 snDNA proximal site on the *A. affinis* Z chromosome (or
20 adjacent to it) could be involved in a DNA double-strand break. Thus, the break triggered
21 the fission and the pericentric inversions repositioned U4 snDNA sites on W₁ and W₂
22 chromosomes.

23 *Parodon* lineage shares a probable homeologous chromosomal U4 snDNA site
24 in chromosome 1. A single U4 snDNA site was also detected in *H. multifasciatus* (Soares
25 et al., 2021). Specifically in *P. hilarii*, the U4 snDNA sequences were also dispersed along
26 the W-specific sex chromosome region. The U4 snDNA dispersion on the *P. hilarii* W
27 chromosome suggests the occurrence of a mechanism of pseudogenic snRNA copies
28 invasion mediated by retrotransposition. This dispersion helped on W sex chromosome
29 differentiation in *P. hilarii*. In addition, no snDNAs signals were identified on the proto-sex
30 chromosome pair proposed by Schemberger et al. (2011) in *P. nasus*. This condition
31 reinforces the proposal that the accumulation of snDNA copies on sex chromosomes in

1 *Apareiodon* sp. and *P. hilarii* were species-specific and acted in each lineage
2 differentiation. One of the critical features of genomes and sex chromosome differentiation
3 is the accumulation of repetitive DNAs (Charlesworth et al., 1994; Charlesworth, 2017). In
4 Parodontidae, the expansion of repetitive DNA units on the W chromosome from an
5 ancestral homomorphic pair is the process by which most species appear to have
6 undergone sex chromosome differentiation (Bellafronte et al., 2011; Schemberger et al.,
7 2011; 2014; 2016; 2019).

8 The U5 and U6 snDNAs presented single sites demonstrating a shared location
9 in Parodontidae karyotypes. Chromosomal arranges carrying 5S rDNA + U1, U2 and U5
10 snDNAs or U2 + U5 snDNAs were identified in fish (Manchado et al., 2006; Merlo et al.,
11 2010, 2012a, 2013), corroborating transpositions occurrence and a diversified scenario of
12 the karyotypic U snDNA evolution in Teleostei. On the other hand, the U6 snDNA mapped
13 by Soares et al. (2021) was located in a single chromosomal pair in *H. multifasciatus*, as
14 observed in Parodontidae species, suggesting their conservative chromosomal location
15 compared to other U snRNAs.

16 According to Marz et al. (2008), snRNAs behave like transposable elements, as
17 they barely demonstrate syntenic positions when evaluating their flanking genes in
18 different genomes. snRNAs could appear associated with rRNAs genes (Manchado et al.,
19 2006; Úbeda-Manzanaro et al., 2010; Scacchetti et al., 2015; Silva et al., 2015; Yano et
20 al., 2017, 2020; Malimpensa et al., 2018, 2020; Piscor et al., 2018), in syntenic arranges
21 of snRNAs (Manchado et al., 2006; Marz et al., 2008; Merlo et al., 2010, 2012a, 2013;
22 García-Souto et al., 2015) or associated to the histone cluster (Malimpensa et al., 2018).
23 However, these clusters are not conserved over longer evolutionary timescales (Marz et
24 al., 2008). The rDNAs were previously mapped in the five species Parodontidae here
25 analyzed (Bellafronte et al., 2005, 2011; Vicari et al., 2006b). In these species, all U
26 snRNA genes were non-syntenic one each other, neither to rRNAs. Furthermore, the
27 genomic analyzes did not show syntenic arranges between rDNA and U snDNA in
28 *Apareiodon* sp.

29 In Parodontidae, chromosomal diversification is proposed to be triggered in
30 repetitive DNA blocks, promoting chromosomal changes (Schemberger et al., 2011, 2014;
31 Ziemniczak et al., 2014; Traldi et al., 2016, 2019; Nascimento et al., 2018; Nirchio et al.,

1 2021). Similarly, snDNAs in situ localization data also demonstrated chromosomal
2 reshuffle events on the Parodontidae karyotype evolution.

3

4 **Conclusion**

5

6 Our data allowed the recognition of nine snRNA gene families belonging to the major and
7 minor spliceosomes and the identification of their functional copies in the *Apareiodon* sp.
8 genome. The data also showed a large number of defective/pseudogenes snRNAs
9 copies, mainly for the *U1*, *U2*, and *U6 snRNAs*. A retrotransposition-mediated mechanism
10 of the defective snDNA units probably scattered these copies on the genome. The in situ
11 localization of the *U1*, *U2*, *U4*, *U5*, and *U6* snDNAs in the chromosomes of Parodontidae
12 species demonstrated some specific chromosomal remodeling in the lineages. The *U2*
13 and *U4* snDNAs were involved in the sex chromosomes differentiation of *Apareiodon* sp.
14 and *P. hilarii*, respectively. In addition, the identification of the homeologous *U4* snDNA
15 sites on chromosome 1 in Parodontidae species and *Z*, *W₁* and *W₂* chromosomes of *A.*
16 *affinis* adds new data to the proposal of a reciprocal translocation between chromosome
17 1 and proto sex chromosome on the origin of the multiple sex chromosome system.

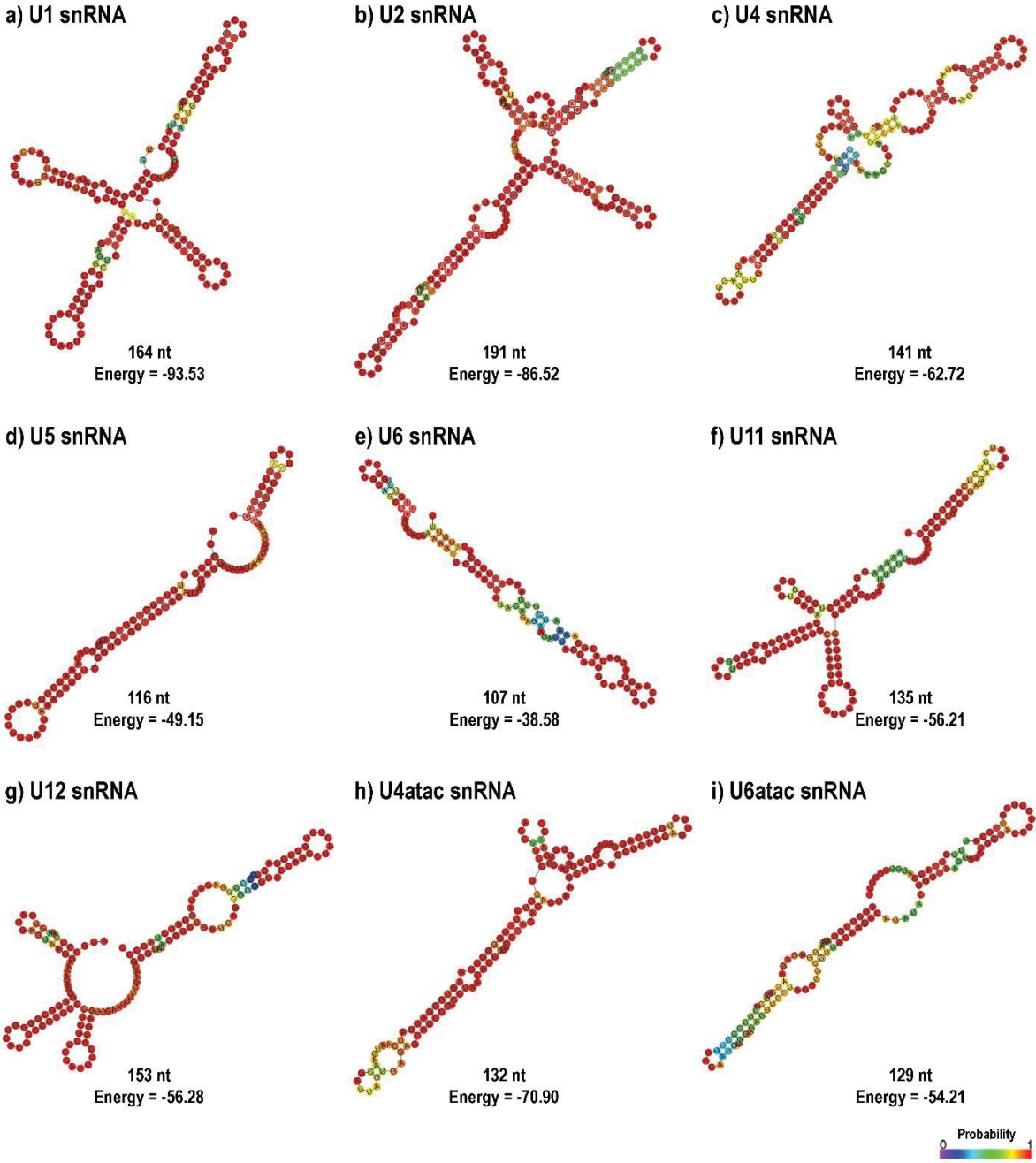
18

19 **References**

20

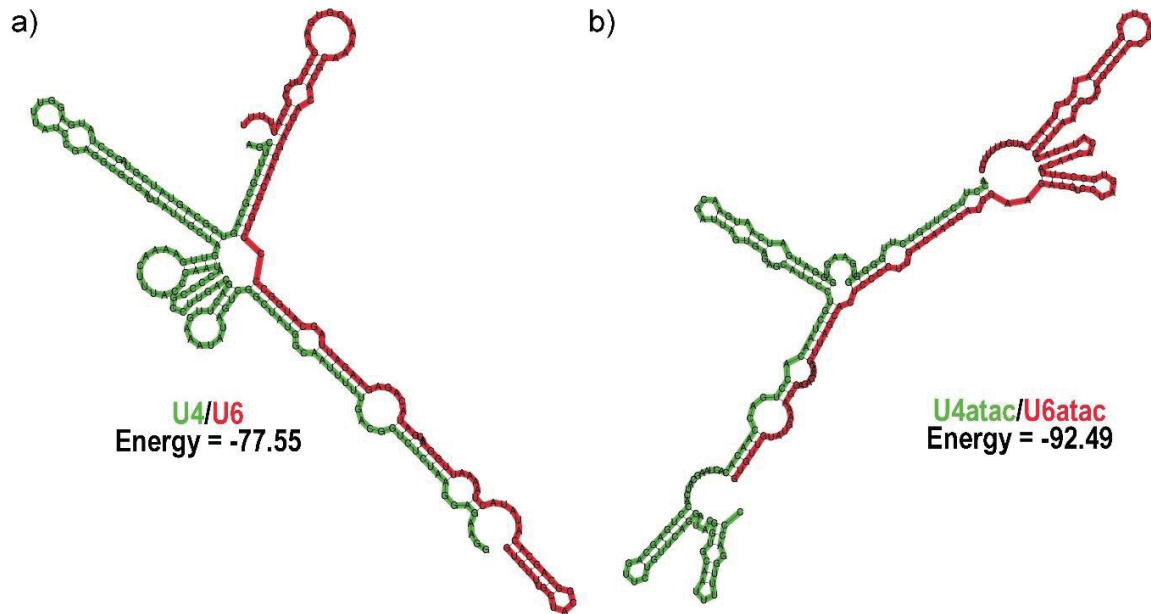
21 As referências deste capítulo encontram-se na seção referências ao final da tese.

1 **Figure 1** – Prediction of the secondary structures of U snRNAs analyzed in the *Apareiodon* sp. genome.



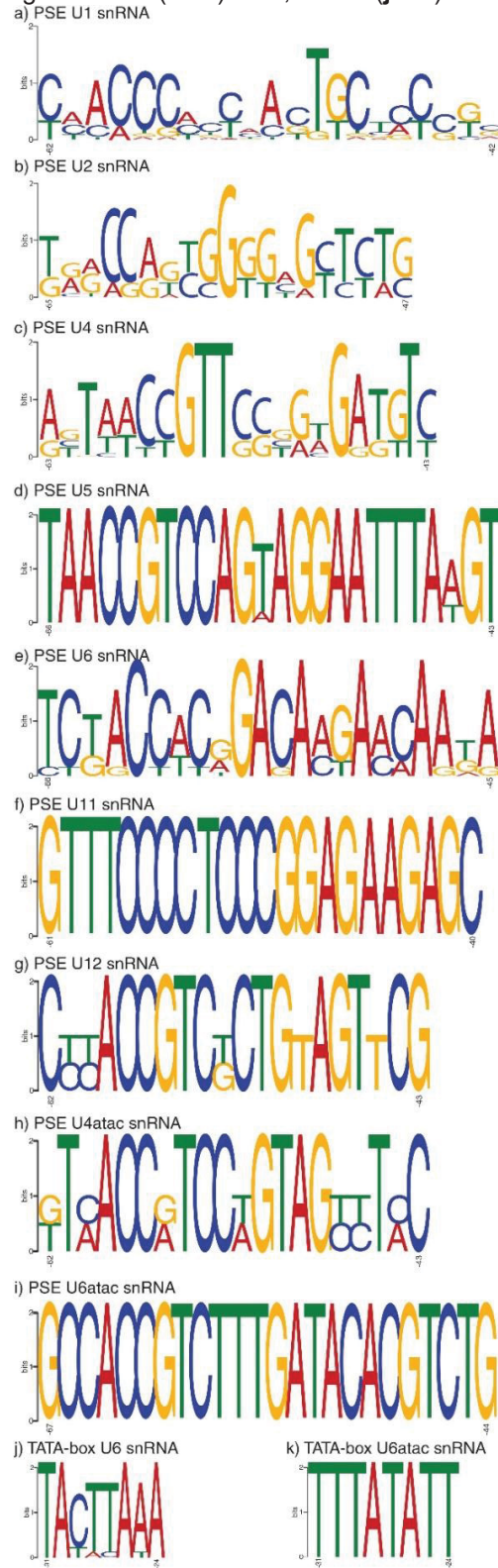
2

- 1 **Figure 2** – Prediction of the *U4/U6* and *U4atac/U6atac* snRNAs complexed structures of the *Apareiodon*
2 sp. In (a) *U4/U6* snRNAs complexed structure; and in (b) *U4atac/U6atac* snRNAs complexed structure.

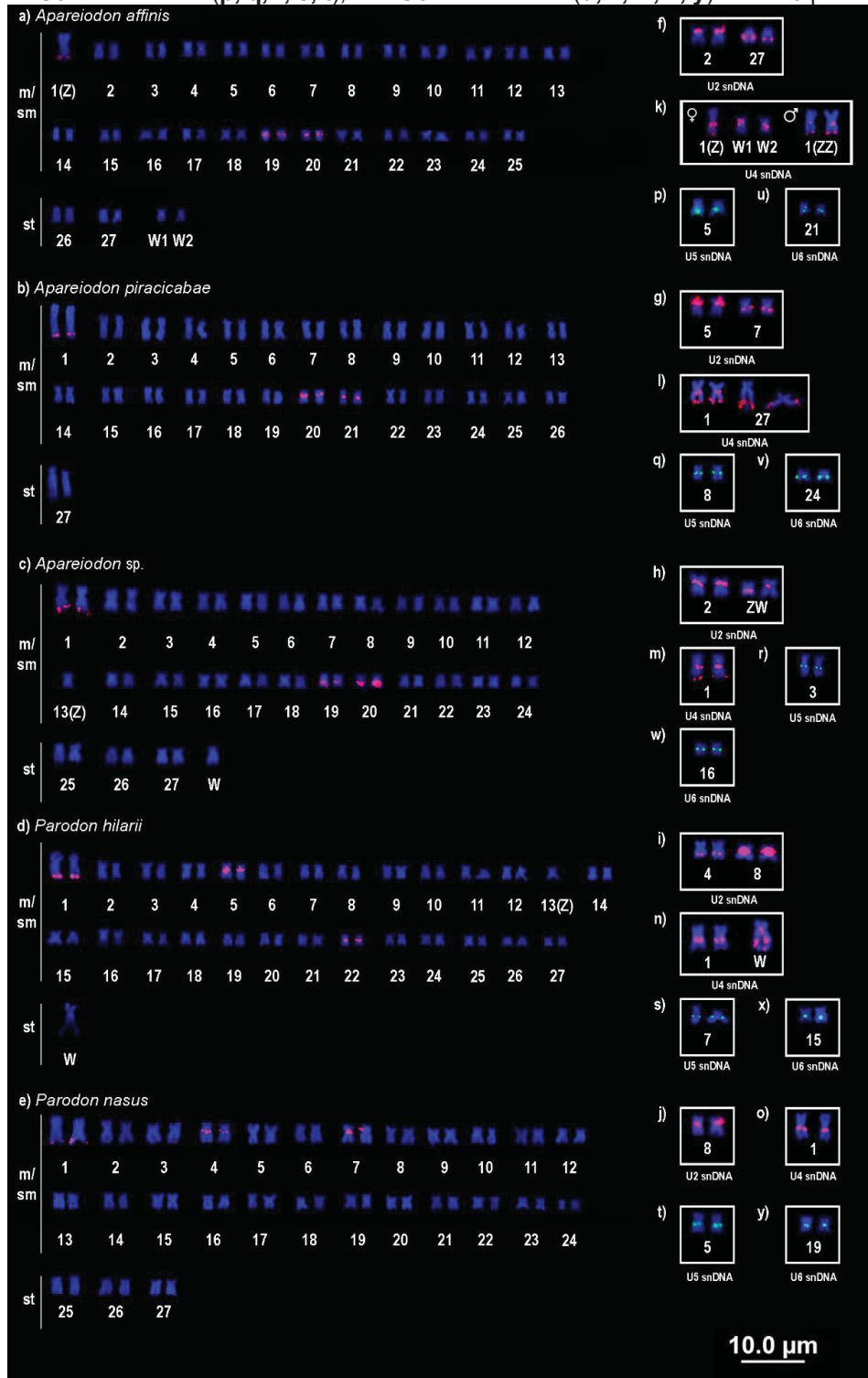


3

1 **Figure 3** – Identification of the consensus regulatory sequences upstream to snRNAs transcribed regions
 2 in *Apareiodon* sp. genome. In (a – i) PSE; and in (j – k) TATA-box consensus.



1 **Figure 4** – In situ localization of the snDNAs sites on the Parodontidae species karyotypes. In (a, b, c, d,
 2 **e**), karyotypes of the *Apareiodon affinis*, *Apareiodon piracicabae*, *Apareiodon* sp., *Parodon hilarii*, and
 3 *Parodon nasus* species, respectively, submitted to FISH using U1 snDNA probe. In the boxes for each
 4 karyotype, respective chromosome bearing U2 snDNA sites (f, g, h, i, j); U4 snDNA sites (k, l, m, n, o);
 5 U5 snDNA sites (p, q, r, s, t); and U6 snDNA sites (u, v, w, x, y). Bar: 10 µm.



1 **Supplementary table 1** – Primers set designed and used to confirm snRNAs in *Apareiodon* sp. genome.

Primer	Sequence (5' – 3')
<i>U1 snRNA</i>	Fwd: TACTACCCTGGCAGGAGTGATAC Rev: AAAGCGTGAGCGCATTG
<i>U2 snRNA</i>	Fwd: CTTCTCGGCCTTTTGGCTAAGAT Rev: TCCCGGAAGTACTGCAA
<i>U4 snRNA</i>	Fwd: CAGTGGCAGTATCGTAGCCTAT Rev: CCAGTCTCCTTAGAAACCGTCAAA
<i>U5 snRNA</i>	Fwd: ACACTCTTGTTTCTTTCAGATCGTA Rev: CCGTAAAGCAGGGCATCAA
<i>U6 snRNA</i>	Fwd: GCTACGGCAGCACATATACTAAAAT Rev: AAAAATGAGGAACGCTTCACG
<i>U11 snRNA</i>	Fwd: AAAAGGGCATCTGTGATGATT Rev: CGTCGAGACGACTGATGAT
<i>U12 snRNA</i>	Fwd: GCCTTAAACTGATGAGTAAGGAAA Rev: CAGCAGAGTAGGCTGGTCA
<i>U4atac snRNA</i>	Fwd: ACYTTCCCTTGTCWTGGGGTGR Rev: GCTTCCAAAAATTGCACCCTT
<i>U6atac snRNA</i>	Fwd: TGTTGTATGAAAGGAGAGAAGGT Rev: TGGTTAGAGATGCCACGAA

2

1 **Supplementary table 2** – Parodontidae species, sex chromosome systems and samples sites.

Species	Sex chromosome system	River	Hydrographic basin	GPS
<i>Apareiodon affinis</i>	ZZ/ZW ₁ W ₂	Passa-Cinco River	Alto Paraná (La Plata)	22°22'19" S; 47°46'54" W
<i>Apareiodon piracicabae</i>	Without heteromorphic chromosomes	Mogi-Guaçu River	Alto Paraná (La Plata)	21°55'34" S; 47°22'03" W
<i>Apareiodon</i> sp.	ZZ/ZW	Verde River	Alto Paraná (La Plata)	25°05'02" S; 50°05'45" W
<i>Parodon hilarii</i>	ZZ/ZW	Araras stream	São Francisco	20°26'16" S; 45°55'39" W
<i>Parodon nasus</i>	Proto-sex chromosome	Passa-Cinco River	Alto Paraná (La Plata)	22°22'19" S; 47°46'54" W

2

1 **Supplementary table 3** – Characterization of snRNAs amplicons obtained from *Apareiodon* sp.

snRNA	Accession number	Sequence size (bp)	Blastn (similarity)	Rfam	CENSOR (coverage)
U1	MZ645211	157	<i>U1 snRNA of Pygocentrus nattereri</i> (XR_005131566.1) (81.46%)	U1 (RF00003) E-value: 6e-18	-
U2	MZ645212	173	<i>U2 snRNA of Pygocentrus nattereri</i> (XR_005131473.1) (95.95%)	U2 (RF00004) E-value: 5.4e-34	-
U4	MZ645213	138	<i>U4 snRNA of Etheostoma cragini</i> (XR_004657755.1) (98.48%)	U4 (RF00015) E-value: 1.7e-21	-
U5	MZ645214	111	<i>U5 snRNA of Electrophorus electricus</i> (XR_003411487.1) (100%)	U5 (RF00020) E-value: 2.1e-14	-
U6	MZ645215	101	<i>U6 snRNA of Pygocentrus nattereri</i> (XR_005129150.1) (100%)	U6 (RF00026) E-value: 3.5e-25	DNA-8N1_HSal (14-66 pb/0.8600)
U11	MZ645216	123	<i>U11 snRNA of Collossoma macropomum</i> (XR_005006628.1) (92.86%)	U11 (RF00548) E-value: 9.7e-16	-
U12	MZ645217	119	<i>U12 snRNA of Pygocentrus nattereri</i> (XR_005129769.1) (94.96%)	U12 (RF00007) E-value: 2.8e-18	-
U4atac	MZ645218	122	<i>U4atac snRNA of Collossoma macropomum</i> (XR_005008192.1) (92.62%)	U4atac (RF00618) E-value: 4e-12	-
U6atac	MZ645219	114	<i>U6atac snRNA of Pygocentrus nattereri</i> (XR_005131685.1) (99.12%)	U6atac (RF00619) E-value: 8.3e-27	-

2

1 4.3 CAPÍTULO 3

2

3

4 Comparative U2 and U4 snDNA chromosomal mapping in the Neotropical fish genera
5 *Apareiodon* and *Parodon* (Characiformes: Parodontidae)

6

7

8 Artigo formatado para ser submetido ao jornal Cytogenetic and Genome Research

1 **Comparative U2 and U4 snDNA chromosomal mapping in the Neotropical fish**
2 **genera *Apareiodon* and *Parodon* (Characiformes: Parodontidae)**

3

4 Matheus Azambuja^a, Viviane Nogaroto^b, Orlando Moreira-Filho^c and Marcelo Ricardo
5 Vicari^{a,b}

6

7 ^a Programa de Pós-Graduação em Genética, Universidade Federal do Paraná, Curitiba,
8 Paraná, Brazil

9 ^b Departamento de Biologia Estrutural, Molecular e Genética, Universidade Estadual de
10 Ponta Grossa, Paraná, Brazil

11 ^c Departamento de Genética e Evolução, Universidade Federal de São Carlos, São
12 Carlos, São Paulo, Brazil

13

14 Short Title: U2 and U4 snDNAs in Parodontidae

15

16 **Abstract**

17 snDNA gene families are considered good cytogenetics markers for comparative studies
18 since different distribution patterns were found among species. In Parodontidae, a
19 Neotropical fish family known to have different sex chromosome systems, the U2 and U4
20 snDNA sites were described in the differentiation of sex chromosomes in some species.
21 However, few studies evaluated snDNA sites as propulsors of chromosome diversification
22 among closely related species in fishes. Here, we investigated the distribution of U2 and
23 U4 snDNA clusters in the chromosomes of ten populations/species belonging to
24 *Apareiodon* and *Parodon*, aiming to find chromosomal homeologies or diversification. Our
25 data revealed a shared submetacentric chromosome pair carrying the U2 snDNA site, and
26 the homeology of the metacentric pair 1 bearing the U4 snDNA cluster among the
27 populations/species analyzed. *Apareiodon* and *Parodon* karyotypes have been described
28 as similar, with the main differences related to heteromorphic sex chromosome
29 differentiation, location and number of the repetitive DNA sites, and the occurrence of
30 some acrocentrics chromosomes in some populations. Despite the previous analyses
31 demonstrating that U2 and U4 snDNA were involved in sex chromosome differentiation in

1 *Apareiodon* sp., *Apareiodon affinis*, and *Parodon hilarii*, the Parodontidae
2 populations/species here analyzed had conserved snDNA sites among the karyotypes.

3

4 **Keywords:** Comparative cytogenetic; Gene families; FISH; Karyotypes; Repetitive DNA.

5

6 **Introduction**

7

8 Repetitive DNA corresponds to the major fraction of the DNA content of many
9 species [Charlesworth et al., 1994; Biscotti et al., 2015]. It is represented by the tandem
10 repeats (satellites, minisatellites and microsatellites), gene families (e.g., ribosomal DNAs,
11 histones, and small nuclear RNAs), and transposable elements [Sumner et al., 2003;
12 López-Flores and Garrido-Ramos, 2012]. Repetitive sequences were known to be
13 responsible for a large fraction of the karyotype variation observed among species
14 [Biscotti et al., 2015; Takagui et al., 2022; Takki et al., 2022].

15 The gene families' units originated from a common ancestral copy in similar DNA
16 sequences and related functions [Nei and Rooney, 2005]. U small nuclear RNAs
17 (snRNAs) are crucial components of the spliceosome, complex machinery responsible for
18 the splicing of mRNA precursors [Will and Lührmann, 2011]. Multiple proteins and the *U1*,
19 *U2*, *U4*, *U5*, and *U6* snRNAs organize the major spliceosome, the dominant splicing
20 machinery [Patel and Steitz, 2003; Sheth et al., 2006]. Their loci are considered good
21 markers for evolutionary studies among closely related species, for inferring homeology
22 between chromosomes from different lineages, and for tracking the origin and evolution
23 of specific genomic regions, such as B and sex chromosomes [Utsunomia et al., 2014].
24 However, comparative cytogenetic studies applying U snDNA probes are still scarce in
25 fishes [Paim et al., 2018].

26 In general, the studies focus on the *in situ* locations of the U1 and U2 snDNAs,
27 with the single-locus condition being the most diagnosed among fish karyotypes [Cabral-
28 de-Mello et al., 2012; Ponzio et al., 2018; Malimpensa et al., 2018, 2020; Haerter et al.,
29 2022]. Though, multiple chromosome sites [Silva et al., 2015; Piscor et al., 2016; Yano et
30 al., 2020; Azambuja et al., 2022a] or scattered signals across the chromosomes were also
31 observed [Úbeda-Manzanaro et al., 2010; Hatanaka et al., 2018]. In addition to U1 and

1 U2 snDNAs, cytogenetic studies are rare for the U4, U5 and U6 snDNAs [Soares et al.,
2 2021; Azambuja et al., 2022a].

3 Parodontidae is a small Neotropical fish family with 32 valid species [Fricke et al.,
4 2022], grouped into three genera: *Parodon* Valenciennes, 1849, *Saccodon* Kner, 1863
5 and *Apareiodon* Eigenmann, 1916 [Pavanelli, 2003]. The species presents a wide
6 geographic distribution through South America and Panama, except in some coastal
7 Atlantic basins and Patagonia [Pavanelli and Britski, 2003].

8 Cytogenetically, Parodontidae species have a shared diploid number ($2n$) of 54
9 chromosomes [Bellafronte et al., 2011]. Despite the $2n$ conservation, karyotype formulae,
10 number and location of the repetitive DNAs, and presence or absence of heteromorphic
11 sex chromosomes were described in the Parodontidae karyotype evolution [Bellafronte et
12 al., 2011; Schemberger et al., 2011; Ziemniczak et al., 2014; Traldi et al., 2016, 2020;
13 Santos et al., 2019; Nirchio et al., 2021; Azambuja et al., 2022a, b]. In this fish family were
14 described species with no evidence of heteromorphic sex chromosomes, species
15 presenting a proto-sex chromosome pair, and others with ZZ/ZW and ZZ/ZW₁W₂ sex
16 chromosome systems [Schemberger et al., 2011; Bellafronte et al., 2012; Traldi et al.,
17 2016, 2020; Nascimento et al., 2018; Santos et al., 2019; Nirchio et al., 2021].

18 Recently, in addition to molecular characterization of the U snRNA genes, the U1,
19 U2, U4, U5, and U6 snDNAs were mapped in the chromosomes of five Parodontidae
20 species [Azambuja et al., 2022a]. The U2 and U4 snDNAs sites were localized in the Z
21 and W chromosomes and proposed to act in sex chromosome differentiation events in
22 some lineage-specific [Azambuja et al., 2022a]. Thus, this study aimed to investigate the
23 karyotype distribution of the U2 and U4 snDNAs in different populations/species of
24 Parodontidae.

25

26 **Material and Methods**

27

28 Individuals from different populations/species of *Apareiodon* and *Parodon*
29 sampled in different River basins in Brazil (Table 1) were analyzed. Specimens were
30 collected with the authorization of the Chico Mendes Institute for Biodiversity Conservation
31 (ICMbio), System of Authorization and Information about Biodiversity (SISBIO-License

1 No. 15117-3), and National System of Genetic Resource Management and Associated
2 Traditional Knowledge (SISGEN No. AE12D3D). The procedures were approved by the
3 Ethics Committee of Animal Usage of the Universidade Estadual de Ponta Grossa
4 (Process CEUA No. 6/2019) and Biosafety Certification according to Comissão Técnica
5 Nacional de Biossegurança - CTNBio (CQB No. 0063/98).

6 Mitotic chromosomes were obtained from the anterior kidney by the air-drying
7 method, according to Bertollo et al. [2015]. The U2 and U4 snDNAs probes were obtained
8 by Polymerase Chain Reaction (PCR) using digoxigenin-11dUTP (Jena Bioscience, Jena,
9 Germany) from the *Apareiodon* sp. genome, according to Azambuja et al. [2022a].
10 Fluorescence *in situ* hybridization (FISH) was performed following Pinkel et al. [1986], and
11 the hybridization was conducted under stringency conditions as follow: 200 ng of each
12 probe, 50% formamide, 10% dextran sulfate, and 2XSSC – saline-sodium citrate; 16 h of
13 hybridization at 37 °C. Signals were detected using anti-digoxigenin rhodamine Fab
14 fragments (Roche Applied Science, Penzberg, Germany). Chromosomes were counter-
15 stained with 0.2 µg/mL of DAPI in VECTASHIELD mounting medium (Vector Laboratories,
16 Burlingame, USA), and chromosomal preparations were analyzed in an epifluorescence
17 microscope (Leica DM 2000) coupled to a DFC3000 G CCD camera (Leica).
18 Chromosomes were classified based on the arm ratio according to Levan et al. (1964)
19 and arranged in metacentric (m), submetacentric (sm), subtelocentric (st), and acrocentric
20 (a) into the karyotypes.

21

22 **Results**

23

24 All populations/species analyzed presented $2n=54$ chromosomes, distinct
25 karyotype formulas, and ZW sex chromosome system occurrence specifically to *A.*
26 *ibitiensis* and *A. vladii* (Table 2, Fig. 1). FISH with the U2 snDNA probe demonstrates a
27 single cluster located at the proximal region of p arm of a submetacentric in all analyzed
28 populations/species (Table 2, Fig.1). This submetacentric corresponds to the
29 chromosome pair 4 in *A. affinis* populations (Table 2, Fig.1 a, b, c), chromosome pair 3 in
30 *A. ibitiensis*, *A. vittatus*, *A. vladii*, *P. pongoensis*, and *Parodon* sp. (Table 2, Fig. 1d, e, f,
31 g, i, j), and chromosome pair 2 in *P. buckleyi* (Table 2, Fig. 1h).

1 In situ localization of the U4 snDNA showed two chromosomal sites on the
2 metacentric 1, a proximal and another terminal, in all *Apareiodon* species (Table 2, Fig.
3 1a – g – boxes). In *Parodon* species, i.e., *P. buckleyi*, *P. pongoensis*, and *Parodon* sp., a
4 single locus at the proximal region of the metacentric 1 was observed (Table 2, Fig. 1h –
5 j – boxes).

6

7 Discussion

8

9 Parodontidae species present a conserved 2n and structural chromosome
10 rearrangements as the main propulsors of karyotype evolution [Bellafronte et al., 2011;
11 Schemberger et al., 2011; Nascimento et al., 2018; Azambuja et al., 2022a, b]. Gene
12 families mapped in the chromosomes of Parodontidae species, such as rDNAs, histones,
13 and snDNAs, have shown few chromosomal remodeling events, most triggering specific
14 events of differentiation in lineage [Bellafronte et al., 2011; Schemberger et al., 2011;
15 Traldi et al., 2019; Azambuja et al., 2022a]. In this study, the comparative in situ
16 localization of the U2 and U4 snDNA probes in a greater number of representatives
17 showed a better diversification scenario between closely related species in Parodontidae.

18 Comparative studies using U2 snDNA probes in fish show that it is usually located
19 on a probable homeologous chromosome pair [Merlo et al., 2010, 2013; Utsunomia et al.,
20 2014; Scacchetti et al., 2015; Ponzio et al., 2018; Pucci et al., 2018; Dulz et al., 2020;
21 Haerter et al., 2022]. However, multiple loci or scattered signals along all the
22 chromosomes can also be observed [Silva et al., 2015; Piscor et al., 2016; Hatanaka et
23 al., 2018]. In this study, the data demonstrated that a single locus of U2 snDNA in a
24 probably homeologous chromosome site is the predominant condition in *Apareiodon* and
25 *Parodon* species, also described in *P. nasus* [Azambuja et al., 2022a]. Additional U2
26 snDNA sites were described in *A. affinis* (Upper Paraná River population), *A. piracicabae*,
27 *Apareiodon* sp. (Verde river), and *P. hilarii* [Azambuja et al., 2022a]. However, these
28 additional sites do not share a chromosomal location, being lineages specific events of
29 differentiation.

30 Studies with the U4 snDNA in fish are restricted to *Hollandichthys multifasciatus*
31 [Soares et al., 2021] and five species of Parodontidae [Azambuja et al., 2022a]. The in

1 situ localization of the U4 snDNA probes showed a shared site in the q proximal region of
2 chromosome 1 in all analyzed Parodontidae species. In addition, all *Apareiodon* species
3 also showed a q terminal U4 snDNA site in chromosome 1. These sites arrange a
4 consistent karyotype divergence between *Parodon* and *Apareiodon* species. Despite
5 these homeologous sites in *Parodon* and *Apareiodon*, Azambuja et al. [2022a] described
6 extra U4 snDNA sites in *A. affinis*, *A. piracicabae*, and *P. hilarii*, but these sites were
7 differentiated specifically in each lineage.

8 The karyotypes in Parodontidae species are predominantly composed of bi-
9 armed chromosomes [Bellafronte et al., 2011; 2012; Traldi et al., 2016, 2020; Nirchio et
10 al., 2021]. The exception occurs in *A. affinis*, which was considered a cryptic species
11 group [Nascimento et al., 2018]. In *A. affinis*, despite $2n=54$ chromosomes conservation,
12 a population presented a multiple sex chromosome system (ZZ/ZW_1W_2), and others
13 showed a variation of 4 to 16 acrocentrics [Moreira-Filho et al., 1980; Jesus et al., 1999;
14 Jorge and Moreira-Filho, 2000; Nascimento et al., 2018]. In this study, the representatives
15 of distinct *A. affinis* populations showed homeologous chromosomes bearing U2 and U4
16 snDNA sites, despite the karyotypes with four, ten or sixteen acrocentrics to the Uruguay,
17 Cuiabá, and Paraguay populations, respectively. The U2 and U4 snDNA sites were also
18 mapped in the karyotypes of *A. affinis* with a ZZ/ZW_1W_2 sex chromosome system, i.e.,
19 from the Upper Paraná River population [Azambuja et al., 2022a]. Regarding the U2
20 snDNA, besides the proximal site in the p arm in a submetacentric, the *A. affinis* from the
21 Upper Paraná River population showed an additional site at the terminal region of the st
22 27. Although morphological analyzes have not demonstrated consistent differences
23 among *A. affinis* representatives [Pavanelli, 2003], the cytogenetic and molecular data
24 support the hypothesis of the cryptic species group among Upper Paraná, Cuiabá,
25 Paraguay, and Uruguay populations [Nascimento et al., 2018].

26 The U4 snDNA sites were mapped in the chromosomes Z, W_1 and W_2 of the *A.*
27 *affinis* from the Upper Paraná River population [Azambuja et al., 2022a]. Based on in situ
28 localization of the repetitive DNAs, a mechanism for the origin of the ZZ/ZW_1W_2 sex
29 chromosome system of the *A. affinis* has been suggested [Schemberger et al., 2011]. In
30 the first stage, a reciprocal translocation occurs between an autosomal pair (metacentric
31 pair 1) and the proto-sex chromosome pair (Fig. 2). After that, the W chromosome

1 generated by reciprocal translocation undergoes centric fission followed by pericentric
2 inversions, originating the W_1 and W_2 sex chromosomes (Fig. 2). In this study, the data
3 demonstrated that the U4 snDNA sites are shared in the q arm of chromosome 1 from the
4 Uruguay, Cuiabá, and Paraguay *A. affinis* populations with the chromosome Z (Fig. 2).
5 Additionally, according to Azambuja et al. [2022a], the pericentromeric U4 snDNA sites in
6 the W_1 and W_2 chromosomes are due to repositioning generated by pericentric inversions
7 (Fig. 2). Thus, the homeology of the arm 1q in all *Apareiodon* species reinforces its
8 involvement in the origin of the multiple sex chromosome system in populations of *A.*
9 *affinis* from Upper Paraná. In this proposal, the arm 1p of the chromosome Z in *A. affinis*
10 corresponds to a large portion of proto-sex chromosome involved in reciprocal
11 translocation.

12 Morphological and cytogenetic analyzes demonstrated that *A. ibitiensis*, *A. vladii*,
13 and *Apareiodon* sp. (Verde River) arrange a closely related group in *Apareiodon*, all of
14 them possessing a ZZ/ZW sex chromosome system, and some doubts regarding
15 *Apareiodon* sp. as a valid species [Rosa et al., 2006; Vicari et al., 2006; Bellafronte et al.,
16 2009]. Despite the shared U2 and U4 snDNA sites in the karyotypes, *Apareiodon* sp. has
17 an additional U2 snDNA site in ZW sex chromosomes, which added to other chromosome
18 differences [Schemberger et al., 2011, 2019] corroborates its diversification from the *A.*
19 *ibitiensis* lineage. *Apareiodon piracicabae* and *A. vittatus* also arrange a paired species
20 group in *Apareiodon* [Pavanelli, 2003], neither of both showing evidence of heteromorphic
21 sex chromosomes [Schemberger et al., 2011], and low rates of genetic divergence
22 [Bellafronte et al., 2013; Santos et al., 2019]. However, an extra U2 and U4 snDNA sites
23 found in *A. piracicabae* [Azambuja et al., 2022a] added to other chromosomal differences
24 [Schemberger et al., 2011] reinforces the karyotype diversification between *A. piracicabae*
25 and *A. vittatus*.

26 Among the *Parodon* analyzed species, the conservation of the U2 and U4 snDNA
27 sites localization into karyotypes is more accentuated compared to *Apareiodon* group.
28 *Parodon buckleyi*, *P. pongoensis*, and *Parodon* sp. (Peixe river) shared the U2 and U4
29 snDNA chromosomal sites with *P. nasus*, analyzed by Azambuja et al. [2022a]. In
30 common, they all possess a proto-sex chromosome pair and few chromosome
31 divergences. The exception is described to *P. hilarii*, which has a differentiated ZW sex

1 chromosome system and extra U2 and U4 snDNA sites in its karyotype. In addition to
2 Parodontidae species, snDNA signals on sex chromosomes have been described in
3 *Gymnotus pantanal* [Utsunomia et al., 2014] and *Triportheus albus* [Yano et al., 2017]. In
4 all lineages, the snDNA sequences participated in specific events of sex chromosome
5 differentiation.

6 Previous studies in Parodontidae have demonstrated the role of repetitive DNAs
7 in chromosomal diversification and specific events of sex chromosomes differentiation
8 [Bellafronte et al., 2011; Schemberger et al., 2011, 2014, 2016, 2019; Ziemniczak et al.,
9 2014; Traldi et al., 2016, 2019; Nascimento et al., 2018; Azambuja et al., 2022a, 2022b].
10 In this study, the comparative localization of the U2 and U4 snDNAs sites showed that
11 these gene families are located predominantly to homeologous chromosomes in
12 Parodontidae species. However, the cytogenetic comparison also detected some species-
13 specific events of differentiation involving snDNA sequences, mainly to heteromorphic sex
14 chromosomes.

15

16 **References**

17

18 As referências deste artigo encontram-se na seção referências ao final da tese.

19

20 **Acknowledgement**

21

22 The authors are grateful to ICMBio (Instituto Chico Mendes de Conservação da
23 Biodiversidade) for authorizing the collections samples.

24

25 **Statement of Ethics**

26

27 Fish were collected with authorization of Instituto Chico Mendes para conservação da
28 Biodiversidade (ICMBio – License 15117–2), and procedures were in agreement with the
29 ethics uses of animals according to the Comitê de Experimentação Animal from
30 Universidade Estadual de Ponta Grossa (Process 6/2019).

31

1 **Conflict of Interest Statement**

2

3 The authors have no conflicts of interest to declare.

4

5 **Funding Sources**

6

7 This study was supported by the Fundação Araucária (Fundação Araucária de Apoio ao
8 Desenvolvimento Científico e Tecnológico do Estado do Paraná, grant number: 9/2017),
9 FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, grant number:
10 2015/16661–1), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
11 - Finance Code 001) and CNPq (Conselho Nacional de Desenvolvimento Científico e
12 Tecnológico, grant number: 305142/2019–4).

13

14 **Author Contributions**

15

16 MA, VN, OMF, and MRV conceived the project ideas; MA performed the experiments;
17 MA, VN, OMF, and MRV analyzed data and wrote the paper.

1 **Table 1** – Species of Parodontidae analyzed and sample location.

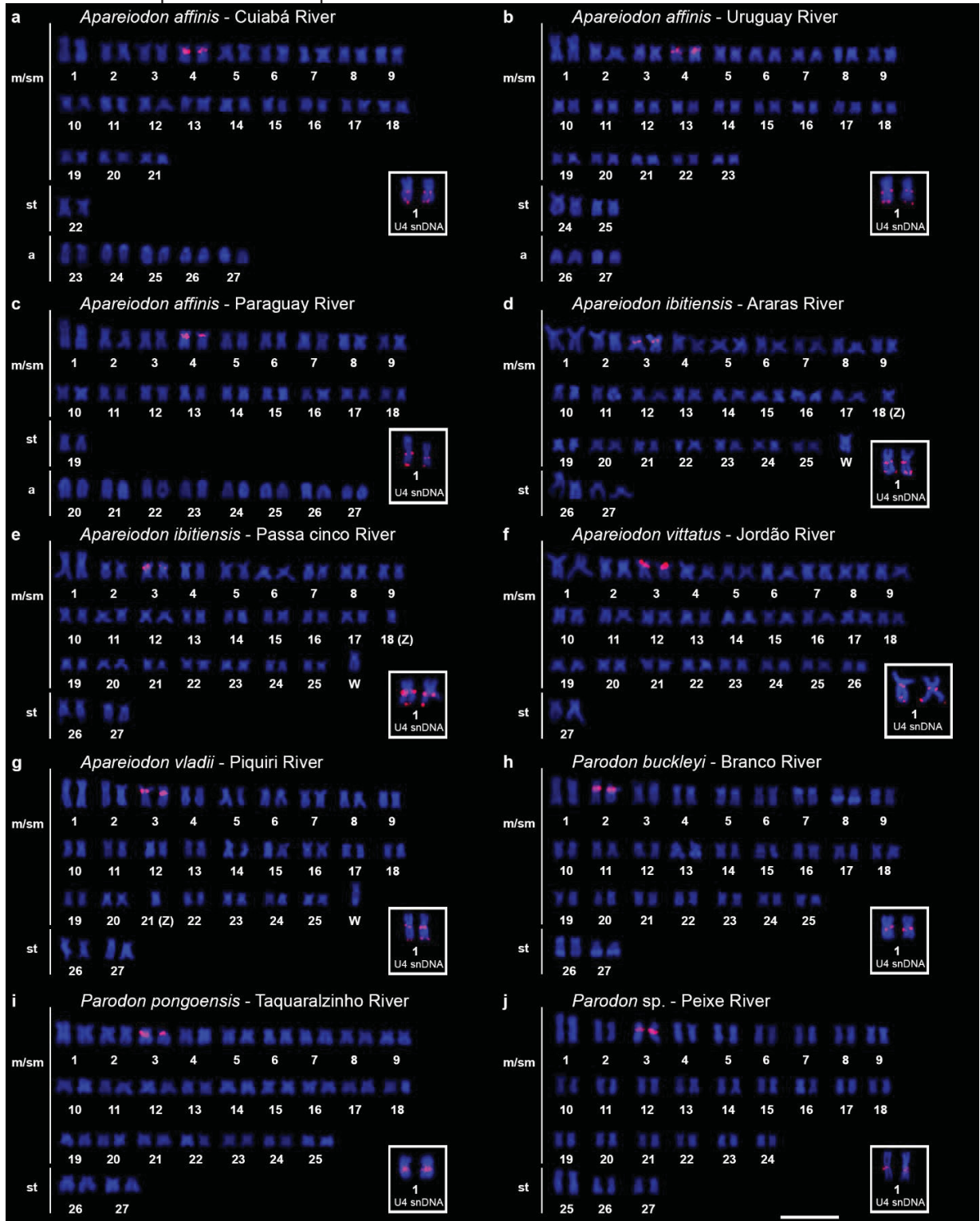
Species	River/Hydrographic basin	Studied specimens	Voucher ID	GPS
<i>Apareiodon affinis</i>	Cuiabá River/Upper Paraguay river basin	5 ♂ 6 ♀	NUP16265	15°34'41" S; 56°09'59" W
<i>Apareiodon affinis</i>	Uruguay River/Upper Uruguay river basin	8 ♂ 9 ♀	NUP16270	27°05'21" S; 53°01'01" W
<i>Apareiodon affinis</i>	Paraguay River/Upper Paraguay river basin	7 ♂ 9 ♀	NUP16264	16°04'31" S; 57°42'10" W
<i>Apareiodon ibitiensis</i>	Araras stream/São Francisco river basin	8 ♂ 9 ♀	MNRJ32741	20°26'16" S; 45°55'39" W
<i>Apareiodon ibitiensis</i>	Passa Cinco River/Upper Paraná river basin	5 ♂ 7 ♀	MNRJ32771	22°22'19" S; 47°46'54" W
<i>Apareiodon vittatus</i>	Jordão River/Iguaçu river basin	5 ♂ 3 ♀	NUP19987	25°35'50.3" S; 51°40'44.6" W
<i>Apareiodon vladii</i>	Piquiri River/Upper Paraná river basin	10 ♂ 10 ♀	NUP3375; NUP3376	25°02'57.6" S; 52°24'22.7" W
<i>Parodon buckleyi</i>	Branco River/Amazon basin	22 ♂ 10 ♀	NUP23214	11°55'51.1" S; 62°09'09.4" W
<i>Parodon pongoensis</i>	Taquaralzinho River/Tocantins-Araguaia basin	4 ♂ 5 ♀	NUP12147	15°53'28" S; 52°14'56" W
<i>Parodon</i> sp.	Peixe River/Amazon basin	8 ♂ 2 ♀	NUP23216	08°39'15.5" S; 55°09'24.3" W

2 MNRJ: Museu Nacional do Rio de Janeiro; NUP: Núcleo de Pesquisas em Limnologia, Ictiologia e
3 Aquicultura (NUPELIA).

Table 2 – *Apareiodon* and *Parodon* cytogenetic data and snDNAs sites. m: metacentric; sm: submetacentric; st: subtelocentric; p: short arm; q: long arm; PS: present study.

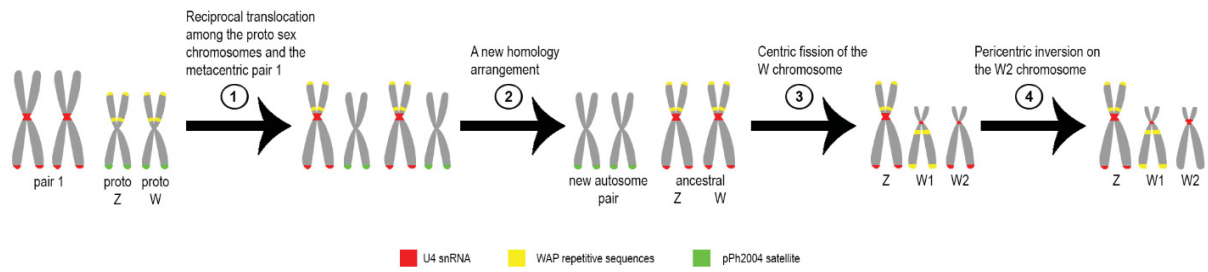
Species	River	2n	Karyotype formula	Sex chromosomes system	U2 snDNA sites	U4 snDNA sites	Reference
<i>Apareiodon affinis</i>	Cuiabá River	54	42m/sm+2st+10a ♂♀	No heteromorphic chromosomes	Proximal sm 4p	Proximal and terminal m 1q	PS
<i>Apareiodon affinis</i>	Uruguay River	54	46m/sm+4st+4a ♂♀	No heteromorphic chromosomes	Proximal sm 4p	Proximal and terminal m 1q	PS
<i>Apareiodon affinis</i>	Paraguay River	54	36m/sm+2st+16a ♂♀	No heteromorphic chromosomes	Proximal sm 4p	Proximal and terminal m 1q	PS
<i>Apareiodon ibitiensis</i>	Araras stream	54	50m/sm+4st ♂♀	ZZ/ZW	Proximal sm 3p	Proximal and terminal m 1q	PS
<i>Apareiodon ibitiensis</i>	Passa Cinco River	54	50m/sm+4st ♂♀	ZZ/ZW	Proximal sm 3p	Proximal and terminal m 1q	PS
<i>Apareiodon vittatus</i>	Jordão River	54	52m/sm+2st ♂♀	No heteromorphic chromosomes	Proximal sm 3p	Proximal and terminal m 1q	PS
<i>Apareiodon vladii</i>	Piquiri River	54	50m/sm+4st ♂♀	ZZ/ZW	Proximal sm 3p	Proximal and terminal m 1q	PS
<i>Parodon buckleyi</i>	Branco River	54	50m/sm+4st ♂♀	Proto-sex chromosome	Proximal sm 2p	Proximal m 1q	PS
<i>Parodon pongoensis</i>	Taquaralzinho River	54	50m/sm+4st ♂♀	Proto-sex chromosome	Proximal sm 3p	Proximal m 1q	PS
<i>Parodon sp.</i>	Peixe River	54	48m/sm+6st ♂♀	Proto-sex chromosome	Proximal sm 3p	Proximal m 1q	PS
<i>Apareiodon affinis</i>	Passa Cinco River	54♂ 55♀	50m/sm+4st ♂ 49m/sm+6st ♀	ZZ/ZW ₁ W ₂	Proximal sm 2; Distal st 27q	Proximal and terminal Zq; Pericentromeric W ₁ W ₂	Azambuja et al. (2022a)
<i>Apareiodon piracicabae</i>	Mogi-Guaçu River	54	52m/sm+2st ♂♀	No heteromorphic chromosomes	Proximal sm 5p; Interstitial sm 7q	Proximal and terminal m 1q; Distal st 27q	Azambuja et al. (2022a)
<i>Apareiodon sp.</i>	Verde River	54	48m/sm+6st ♂ 47m/sm+7st ♀	ZZ/ZW	Centromeric m 2; Proximal ZW	Proximal and terminal m 1q	Azambuja et al. (2022a)
<i>Parodon hilarii</i>	Araras stream	54	54m/sm ♂ 53m/sm+1st ♀	ZZ/ZW	Distal sm 4q; Centromeric m 8	Proximal m 1q; Dispersed Wq	Azambuja et al. (2022a)
<i>Parodon nasus</i>	Passa Cinco River	54	48m/sm+6st ♂♀	Proto-sex chromosome	Proximal sm 8	Proximal m 1q	Azambuja et al. (2022a)

1 **Figure 1** – Karyotypes of *Apareiodon affinis* populations (a, b, c), *A. ibitiensis* populations (d, e), *A. vittatus*
 2 (f), *A. vladii* (g), *Parodon buckleyi* (h), *P. pongoensis* (i), and *Parodon* sp. (j) submitted to FISH using U2
 3 and U4 snDNAs probes. Bar = 10 µm.



4
5

1 **Figure 2** – Origin and differentiation of the ZZ/ZW₁W₂ sex chromosomes system in *Apareiodon affinis* from
 2 the Upper Paraná River.



3
4

1 5 CONSIDERAÇÕES FINAIS

2

3 A obtenção de dados de sequenciamento de nova geração e a montagem do
4 genoma de *Apareiodon* sp. (Rio Verde) permitiram a utilização de novas ferramentas e
5 metodologias para o entendimento da porção repetitiva do genoma desta espécie. Os
6 primeiros estudos com o genoma de *Apareiodon* sp. focaram na caracterização geral dos
7 elementos repetitivos e no entendimento da origem e diferenciação do cromossomo W.

8 Neste estudo, a prospecção em larga escala de MITEs foi realizada pela primeira
9 vez no genoma de uma espécie de peixe. Os dados demonstraram que a maior parte das
10 sequências observadas foram originadas a partir de TEs das superfamílias *Tc-1/Mariner*
11 e *hAT*. A localização das MITEs próximas ou dentro genes, bem como, suas identidades
12 com ncRNA, os qualificam para processos de co-opção molecular, os quais serão
13 integrados com estudos de RNAseq. A similaridade com famílias de pre-miRNAs e com
14 miRNAs maduros de outros peixes, demonstraram a provável co-opção molecular de
15 cópias MITEs para ncRNAs. A obtenção de dados de RNAseq, juntamente a análises de
16 identificação de miRNAs e de alvos para estes devem validar as informações
17 encontradas.

18 Neste estudo ainda foi realizada a predição e caracterização dos snRNAs
19 envolvidos nos processos de *splicing* pelos dois tipos de spliceossomos. A identificação
20 das sequências promotoras PSE e TATA-box, e a correta organização dos snRNAs em
21 estruturas secundárias demonstraram cópias putativamente funcionais no genoma de
22 *Apareiodon* sp. Em outra via, inúmeras cópias defectivas e pseudogênicas foram
23 encontradas, estas cópias provavelmente foram originadas a partir de eventos de
24 retrotransposição destes snRNAs.

25 As análises de localização *in situ* demonstraram diferentes cenários para os
26 snDNAs U1, U2, U4, U5 e U6 nas espécies dos gêneros *Apareiodon* e *Parodon*.
27 Cromossomos homeólogos portadores destas sequências foram propostos, com a
28 identificação de eventos independentes de dispersão para alguns dos snDNAs. A
29 homeologia do par cromossômico 1, portador do snDNA U4, revelou o envolvimento
30 deste par na origem do sistema de cromossomos sexuais múltiplo (ZZ/ZW₁W₂) na
31 população de *A. affinis* do Alto Rio Paraná.

1 **6 REFERÊNCIAS**

- 2
- 3 ALBUQUERQUE, N. R. M.; EBERT, D.; HAAG, K. L. Transposable element abundance
4 correlates with mode of transmission in microsporidian parasites. **Mobile DNA**, v. 11, n.
5 19, 2020. <https://doi.org/10.1186/s13100-020-00218-8>
- 6 ALONGE, M.; SOYK, S.; RAMAKRISHNAN, S.; WANG, X.; GOODWIN, S.; SEDLAZECK,
7 F. J.; LIPPMAN, Z. B.; SCHATZ, M. C. RaGOO: fast and accurate reference-guided
8 scaffolding of draft genomes. **Genome Biology**, v. 20, n. 224, 2019.
9 <https://doi.org/10.1186/s13059-019-1829-6>
- 10 ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local
11 alignment search tool. **J Mol Biol**, v. 215, p. 403–410, 1990.
12 [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 13 AMARAL, P. P.; DINGER, M. E.; MERCER, T. R.; MATTICK, J. S. The Eukaryotic
14 Genome as an RNA Machine. **Science**, v. 319, p. 1787 – 1789, 2008.
15 <https://doi.org/10.1126/science.1155472>
- 16 ANJOS, A.; RUIZ-RUANO, F. J.; CAMACHO, J. P. M.; LORETO, V.; CABRERO, J.; dE
17 SOUZA, M. J.; CABRAL-DE-MELLO, D. C. U1 snDNA clusters in grasshoppers:
18 chromosomal dynamics and genomic organization. **Heredity**, v. 114, p. 207–219, 2015.
19 <https://doi.org/10.1038/hdy.2014.87>
- 20 ARAYA-JAIME, C.; LAM, N.; PINTO, I. V.; MÉNDEZ, M. A.; ITURRA, P. Chromosomal
21 organization of four classes of repetitive DNA sequences in killifish *Orestias ascotanensis*
22 Parenti, 1984 (Cyprinodontiformes, Cyprinodontidae). **Comp Cytogenet**, v. 11, p. 463–
23 475, 2017. <https://doi.org/10.3897/CompCytogen.v11i3.11729>
- 24 ARAVIN, A. A.; HANNON, G. J.; BRENNECKE, J. The piwi-piRNA pathway provides an
25 adaptive defense in the transposon arms race. **Science**, v.318, p. 761–764, 2007.
26 <https://doi.org/10.1126/science.1146484>
- 27 ARIEL, F. D.; MANAVELLA, P. A. When junk DNA turns functional: transposon-derived
28 non-coding RNAs in plants. **J Exp Botany**, v. 72, p. 4132–4143, 2021.
29 <https://doi.org/10.1093/jxb/erab073>
- 30 AZAMBUJA, M.; SCHEMBERGER, M. O.; NOGAROTO, V.; MOREIRA-FILHO, O.;
31 MARTINS, C.; VICARI, M. R. Major and minor U small nuclear RNAs genes
32 characterization in a neotropical fish genome: Chromosomal remodeling and repeat units
33 dispersion in Parodontidae. **Gene**, v. 826, n. 146459, 2022a.
34 <https://doi.org/10.1016/j.gene.2022.146459>
- 35 AZAMBUJA, M.; MARCONDES, D. S.; NOGAROTO, V.; MOREIRA-FILHO, O.; VICARI,
36 M. R. Population structuration and chromosomal features homogeneity in *Parodon nasus*

- 1 (Characiformes: Parodontidae): A comparison between Loer and Upper Paraná River
2 representatives. **Neotrop Ichtyol**, v. 20, n. e210162, 2022b. [https://doi.org/10.1590/1982-](https://doi.org/10.1590/1982-0224-2021-0162)
3 0224-2021-0162
- 4 BAILEY, T. L.; BODEN, M.; BUSKE, F. A.; FRITH, M.; GRANT, C. E.; CLEMENTI, L.;
5 REN, J.; LI, W. W.; NOBLE, W. S. MEME SUITE: tools for motif discovery and searching.
6 **Nucleic Acids Res**, v. 37, p. W202–W208, 2009. <https://doi.org/10.1093/nar/gkp335>
- 7 BAO, W.; JURKA, J. DNA transposons from the *Harpegnathos saltator* genome. **Rebase**
8 **Reports**, v. 14, p. 833–833, 2014.
- 9 BARMAN, A. S.; SINGH, M.; SINGH, R. K.; LAL, K. K. Evidence of birth-and-death
10 evolution of 5S rRNA gene in *Channa* species (Teleostei, Perciformes). **Genetica**, v. 144,
11 p. 723–732, 2016. <https://doi.org/10.1007/s10709-016-9938-6>
- 12 BARROS, A. V.; WOLSKI, M. A. V.; NOGAROTO, V.; ALMEIDA, M. C.; MOREIRA-
13 FILHO, O.; VICARI, M. R. Fragile Sites, dysfunctional telomere and chromosome fusions:
14 What is 5S rDNA role?. **Gene**, v. 608, p. 20–27, 2017.
15 <https://doi.org/10.1016/j.gene.2017.01.013>
- 16 BELLAFRONTÉ, E.; MARGARIDO, V. P.; MOREIRA-FILHO, O. Cytotaxonomy of
17 *Parodon nasus* and *Parodon tortuosus* (Pisces, Characiformes). A case of synonymy
18 confirmed by cytogenetic analyses. **Genet Mol Biol**, v. 28, p. 710–716, 2005.
19 <https://doi.org/10.1590/S1415-47572005000500010>
- 20 BELLAFRONTÉ, E.; VICARI, M. R.; ARTONI, R. F.; MARGARIDO, V. P.;
21 MOREIRAFILHO, O. Differentiated ZZ/ZW sex chromosomes in *Apareiodon ibitiensis*
22 (Teleostei, Parodontidae): citotaxonomy and biogeography. **J Fish Biol**, v. 75, p. 2313–
23 2325, 2009. <https://doi.org/10.1111/j.1095-8649.2009.02488.x>
- 24 BELLAFRONTÉ, E.; SCHEMBERGER, M. O.; MOREIRA-FILHO, O.; ALMEIDA, M. C.;
25 ARTONI, R. F.; MARGARIDO, V. P.; VICARI, M. R. Chromosomal markers in
26 Parodontidae: an analysis of new and reviewed data with phylogenetic inferences. **Rev**
27 **Fish Biol Fish**, v. 21, p. 559–570, 2011. <https://doi.org/10.1007/s11160-010-9177-3>
- 28 BELLAFRONTÉ, E.; SCHEMBERGER, M. O.; ARTONI, R. F.; MOREIRA-FILHO, O.;
29 VICARI, M. R. Sex chromosome system ZZ/ZW in *Apareiodon hasemani* Eigenmann,
30 1916 (Characiformes, Parodontidae) and a derived chromosomal region. **Genet Mol Biol**,
31 v. 35, p. 770–776, 2012. <https://doi.org/10.1590/S1415-47572012005000077>
- 32 BERNHART, S. H.; TAFER, H.; MÜCKSTEIN, U.; FLAMM, C.; STADLER, P. F.;
33 HOFACKER, I. L. Partition function and base pairing probabilities of RNA heterodimers.
34 **Algorithms Mol Biol**, v. 1, n. 3, 2006. <https://doi.org/10.1186/1748-7188-1-3>

- 1 BERNHART, S. H.; HOFACKER, I. L.; WILL, S.; GRUBER, A. R.; STADLER, P. F.
2 RNAalifold: improved consensus structure prediction for RNA alignments. **BMC**
3 **Bioinformatics**, v. 9, n. 474, 2008. <https://doi.org/10.1186/1471-2105-9-474>
- 4 BERTOLLO, L. A. C.; CIOFFI, M. B.; MOREIRA-FILHO, O. Direct chromosome
5 preparation from freshwater teleost fishes. *In* Fish Cytogenetic Techniques
6 (Chondrichthyans and Teleosts): OZOUF-COSTAZ, C.; PISANO, E.; FORESTI, F.;
7 ALMEIDA TOLEDO, L. F., Eds.; **CRC Press**: Boca Raton, FL, USA, p. 21–26, 2015
- 8 BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. Repetitive DNA in eukaryotic
9 genomes. **Chromosome Res**, v. 23, p. 415–420, 2015. [https://doi.org/10.1007/s10577-](https://doi.org/10.1007/s10577-015-9499-z)
10 [015-9499-z](https://doi.org/10.1007/s10577-015-9499-z)
- 11 BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina
12 sequence data. **Bioinformatics**, v. 30, p. 2114–2120, 2014.
13 <https://doi.org/10.1093/bioinformatics/btu170>
- 14 BORCHERT, G. M.; LANIER, W.; DAVIDSON, B. L. RNA polymerase III transcribes
15 human microRNAs. **Nature Structural & Molecular Biology**, v. 13, p. 1097 – 1101, 2006.
16 <https://doi.org/10.1038/nsmb1167>
- 17 BRÛNA, T.; HOFF, K. J.; LOMSADZE, A.; STANKE, M.; BORODOVSKY, M. BRAKER2:
18 automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS
19 supported by a protein database. **NAR Genomics and Bioinformatics**, v. 3, n. lqaa108,
20 2021. <https://doi.org/10.1093/nargab/lqaa108>
- 21 BUREAU, T. E.; WESSLER, S. R.; Tourist: a large family of small inverted repeat
22 elements frequently associated with maize genes. **Plant Cell**, v. 4, p. 1283–1294, 1992.
23 <https://doi.org/10.1105/tpc.4.10.1283>
- 24 BURGE, C. B.; PADGETT, R. A.; SHARP, P. A. Evolutionary fates and origins of U12-
25 type introns. **Mol Cell**, v. 2, p. 773–785, 1998. [https://doi.org/10.1016/S1097-](https://doi.org/10.1016/S1097-2765(00)80292-0)
26 [2765\(00\)80292-0](https://doi.org/10.1016/S1097-2765(00)80292-0)
- 27 CABRAL-DE-MELLO, D. C.; VALENTE, G. T.; NAKAJIMA, R. T.; MARTINS, C. Genomic
28 organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish,
29 with an emphasis in *Oreochromis niloticus*. **Chromosome Res**, v. 20, p. 279–292, 2012.
30 <https://doi.org/10.1007/s10577-011-9271-y>
- 31 CAO, L.; ZHAO, C.; WANG, C.; QIN, H.; QIN, Q.; TAO, M.; ZHANG, C.; ZHAO, R.; LIU,
32 S. (2020) Evolutionary dynamics of 18S and 5S rDNA in autotriploid *Carassius auratus*.
33 **Gene**, v. 737, n. 144433, 2020. <https://doi.org/10.1016/j.gene.2020.144433>
- 34 CAVALCANTE, M. G.; BASTOS, C. E. M. C.; NAGAMACHI, C. Y.; PIECZARKA, J. C.;
35 VICARI, M. R.; NORONHA, R. C. R. Physical mapping of repetitive DNA suggests 2n

- 1 reduction in Amazon turtles *Podocnemis* (Testudines: Podocnemididae). **PLoS One**, v.
2 13, n. e0197536, 2018. <https://doi.org/10.1371/journal.pone.0197536>
- 3 CAZAUX, B.; CATALAN, J.; VEYRUNES, F.; DOUZERY, E. J.; BRITTON-DAVIDIAN, J.
4 Are ribosomal DNA clusters rearrangement hotspots?: a case study in the genus *Mus*
5 (Rodentia, Muridae). **BMC Evol Biol**, v. 11, n. 124, 2011. [https://doi.org/10.1186/1471-](https://doi.org/10.1186/1471-2148-11-124)
6 2148-11-124
- 7 CECH, T. R.; STEITZ, J. A. The noncoding RNA revolution – trashing old rules to forge
8 new ones. **Cell**, v. 157, p. 77–94, 2014. <https://doi.org/10.1016/j.cell.2014.03.008>
- 9 CENTOFANTE, L.; BERTOLLO, L. A. C.; MOREIRA-FILHO, O. A ZZ/ZW sex
10 chromosomes system in a new species of the genus *Parodon* (Pisces, Parodontidae).
11 **Caryologia**, v. 55, p. 139–150, 2002. <https://doi.org/10.1080/00087114.2002.10589270>
- 12 CHAIRI, H.; GONZALES, L. R. Structure and organization of the Engraulidae family U2
13 snRNA: an evolutionary model gene?. **J Mol Evol**, v. 80, p. 209–218, 2015.
14 <https://doi.org/10.1007/s00239-015-9674-z>
- 15 CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of
16 repetitive DNA in eukaryotes. **Nature**, v. 371, p. 215–220, 1994.
17 <https://doi.org/10.1038/371215a0>
- 18 CHARLESWORTH, D. Evolution of recombination rates between sex chromosomes.
19 **Philos Trans R Soc Lond B Biol Sci**, v. 372, n. 20160456, 2017.
20 <https://doi.org/10.1098/rstb.2016.0456>CIGANDA, M.; WILLIAMS, N. Eukaryotic 5S rRNA
21 biogenesis. **Wiley Interdiscip Rev RNA**, v. 2, p. 523–533, 2011.
22 <https://doi.org/10.1002/wrna.74>
- 23 CHEN, J.; HU, Q.; LU, C.; KUANG, H. Evolutionary genomics of miniature inverted-repeat
24 transposable elements (MITEs) in plants. *In*: PONTAROTTI, P. (Eds) Evolutionary
25 Biology: Genome evolution, speciation, coevolution and origin of life. Springer, p. 157–
26 168, 2014a. https://doi.org/10.1007/978-3-319-07623-2_7
- 27 CHEN, J.; HU, Q.; ZHANG, Y.; LU, C.; KUANG, H. P-MITE: a database for plant miniature
28 inverted-repeat transposable elements. **Nucleic Acids Res**, v. 42, p. D1176–D1181,
29 2014b. <https://doi.org/10.1093/nar/gkt1000>
- 30 CHEN, Q.; MENG, X.; LIAO, Q.; CHEN, M. Versatile interactions and bioinformatics
31 analysis of noncoding RNAs. **Brief Bioinform**, v. 20, p. 1781–1794, 2019.
32 <https://doi.org/10.1093/bib/bby050>
- 33 CHÉNAIS, B.; CARUSO, A.; HIARD, S.; CASSE, N. The impact of transposable elements
34 on eukaryotic genomes: From genome size increase to genetic adaptation to stressful
35 environments. **Gene**, v. 509, p. 7– 5, 2012. <https://doi.org/10.1016/j.gene.2012.07.042>

- 1 COLLINS, L.; PENNY, D. Complex Spliceosomal Organization Ancestral to Extant
2 Eukaryotes. **Mol Biol Evol**, v. 22, p. 1053–1066, 2005.
3 <https://doi.org/10.1093/molbev/msi091>
- 4 CRESCENTE, J. M.; ZAVALLO, D.; HELGUERA, M.; VANZETTI, L. S. MITE Tracker: an
5 accurate approach to identify miniature inverted-repeat transposable elements in large
6 genomes. **BMC Bioinformatics**, v. 19, n. 348, 2018. [https://doi.org/10.1186/s12859-018-](https://doi.org/10.1186/s12859-018-2376-y)
7 [2376-y](https://doi.org/10.1186/s12859-018-2376-y)
- 8 CUI, J.; YOU, C.; CHEN, X. The evolution of microRNAs in plants. **Current opinion in**
9 **plant biology**, v. 35, p. 61 – 67, 2017. <https://doi.org/10.1016/j.pbi.2016.11.006>
- 10 DEGRANDI, T. M.; GUNSKI, R. J.; GARNERO, A. V.; OLIVEIRA, E. H. C.;
11 KRETSCHMER, R.; SOUZA, M. S.; BARCELLOS, S. A.; HASS, I. The distribution of 45S
12 rDNA sites in bird chromosomes suggests multiple evolutionary histories. **Genet Mol Biol**,
13 v. 43, n. e20180331, 2020. <https://doi.org/10.1590/1678-4685-GMB-2018-0331>
- 14 DEON, G. A.; GLUGOSKI, L.; VICARI, M. R.; NOGAROTO, V.; SASSI, F. M. C.; CIOFFI,
15 M. B.; LIEHR, T.; BERTOLLO, L. A. C.; MOREIRA-FILHO, O. Highly rearranged
16 karyotypes and multiple sex chromosome systems in armored catfishes from the genus
17 *Harttia* (Teleostei, Siluriformes). **Genes**, v. 11, n. 1366, 2020.
18 <https://doi.org/10.3390/genes11111366>
- 19 DEVESON, I. W.; HARDWICK, S. A.; MERCER, T. R.; MATTICK, J. S. The dimensions,
20 dynamics, and relevance of the mammalian noncoding transcriptome. **Trends Genet**, v.
21 33, p. 464–478, 2017. <https://doi.org/10.1016/j.tig.2017.04.004>
- 22 DOUCET, A. J.; DROC, G.; SIOL, O.; AUDOUX, J.; GILBERT, N. U6 snRNA
23 pseudogenes: markers of retrotransposition dynamics in mammals. **Mol Biol Evol**, v. 32,
24 p. 1815–1832, 2015. <https://doi.org/10.1093/molbev/msv062>
- 25 DRONGITIS, D.; ANIELLO, F.; FUCCI, L.; DONIZETTI, A. Roles of transposable
26 elements in the different layers of gene expression regulation. **Int J Mol Sci**, v. 20, n.
27 5755, 2019. <https://doi.org/10.3390/ijms20225755>
- 28 DULZ, T. A.; AZAMBUJA, M.; NASCIMENTO, V. D.; LORSCHIEDER, C. A.; NOLETO, R.
29 B.; MOREIRA-FILHO, O.; NOGAROTO, V.; DINIZ, D.; AFFONSO, P. R. A. M.; VICARI,
30 M. R. Karyotypic diversification in two *Megaleporinus* species (Characiformes,
31 Anostomidae) inferred from *in situ* localization of repetitive DNA sequences. **Zebrafish**,
32 v. 17, p. 333–341, 2020. <https://doi.org/10.1089/zeb.2020.1918>
- 33 EIRÍN-LÓPEZ, J. M.; REBORDINOS, L.; ROONEY, A. P.; ROZAS, J. The birth-and-death
34 evolution of multigene families revisited. **Genome Dyn**, v. 7, p. 170–196, 2012.
35 <https://doi.org/10.1159/000337119>

- 1 FALCÃO, J. N.; MOREIRA-FILHO, O.; BERTOLLO, L. A. C. An additional chromosome
2 in two fish species. **Brazilian Journal of Genetics**, v. 7, p. 109–118, 1984.
- 3 FATTASH, I.; ROOKE, R.; WONG, A.; LUU, T.; BHARDWAJ, P.; YANG, G. Miniature
4 inverted-repeat transposable elements: discovery, distribution, and activity. **Genome**, v.
5 56, p. 475–786, 2013. <https://doi.org/10.1139/gen-2012-0174>
- 6 FERNÁNDEZ-MEDINA, R. D.; RIBEIRO, J. M. C.; CARARETO, C. M. A.; VELASQUES,
7 L.; STRUCHINER, C. J. Losing identity: structural diversity of transposable elements
8 belonging to different classes in the genome of *Anopheles gambiae*. **BMC Genomics**, v.
9 13, n. 272, 2012. <https://doi.org/10.1186/1471-2164-13-272>
- 10 FESCHOTTE, C. Transposable elements and the evolution of regulatory networks. **Nat**
11 **Rev Genet**, v. 9, p. 397–405, 2008. <https://doi.org/10.1038/nrg2337>
- 12 FESCHOTTE, C.; JIANG, N.; WESSLER, S. Plant transposable elements: where genetics
13 meets genomics. **Nat Rev Genet**, v. 3, p. 329–341. <https://doi.org/10.1038/nrg793>
- 14 FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends**
15 **Genet**, v. 5, p. 103–107, 1989. [https://doi.org/10.1016/0168-9525\(89\)90039-5](https://doi.org/10.1016/0168-9525(89)90039-5)
- 16 FRICKER, R.; ESCHMEYER, W. N.; VAN der LAAN, R. Eschmeyer's catalog of fishes:
17 genera, species, references. Disponível em:
18 <https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>,
19 acesso em 02/05/2022.
- 20 GARCÍA-SOUTO, D.; TRONCOSO, T.; PÉREZ, M.; PASANTES, J. J. Molecular
21 Cytogenetic Analysis of the European Hake *Merluccius merluccius* (Merlucciidae,
22 Gadiformes): U1 and U2 snRNA Gene Clusters Map to the Same Location. **PLoS One**,
23 v. 10, n. e0146150, 2015. <https://doi.org/10.1371/journal.pone.0146150>
- 24 GAZONI, T.; DORIGON, N. S.; SILVA, M. J.; CHOLAK, L. R.; HADDAD, C. F. B.; PARISE-
25 MALTEMPI, P. P. Chromosome mapping of U2 snDNA in species of *Leptodactylus*
26 (Anura, Leptodactylidae). **Cytogenet Genome Res**, v. 161, p. 63–69, 2021.
27 <https://doi.org/10.1159/000515047>
- 28 GLUGOSKI, L.; GIULIANO-CAETANO, L.; MOREIRA-FILHO, O.; VICARI, M. R.;
29 NOGAROTO, V. Co-located *hAT* transposable element and 5S rDNA in an interstitial
30 telomeric sequence suggest the formation of Robertsonian fusion in armored catfish.
31 **Gene**, v. 650, p. 49–54, 2018. <https://doi.org/10.1016/j.gene.2018.01.099>
- 32 GORNUNG, E. Twenty years of physical mapping of major ribosomal RNA genes across
33 the Teleosts: a review of research. **Cytogenet Genome Res**, v. 141, p. 90–102, 2013.
34 <https://doi.org/10.1159/000354832>

- 1 GRACE, C. A.; CARR, M. The evolutionary history of *mariner* elements in stalk-eyed flies
2 reveals the horizontal transfer of transposons from insects into the genome of the
3 cnidarian *Hydra vulgaris*. **PLOS ONE**, v. 15, n. e0235984, 2020.
4 <https://doi.org/10.1371/journal.pone.0235984>
- 5 GREENE, B.; WALKO, R.; HAKE, S. Mutator insertions in an intron of the maize knotted1
6 gene result in dominant suppressible mutations. **Genetics**, v. 138, p. 1275 – 1285, 1994.
7 <https://doi.org/10.1093/genetics/138.4.1275>
- 8 HAERTER, C. A. G.; MARGARIDO, V. P.; BLANCO, D. R.; TRALDI, J. B.; FELDBERG
9 E.; LUI, R. L. Contributions to *Trachelyopterus* (Siluriformes: Auchenipteridae) species
10 diagnosis by cytotaxonomic autapomorphies: from U2 snRNA chromosome
11 polymorphism to rDNA and histone gene synteny. **Org Divers Evol**, 2022.
12 <https://doi.org/10.1007/s13127-022-00560-0>
- 13 HATANAKA, T.; OLIVEIRA, E. A.; RÁB, P.; YANO, C. F.; BERTOLLO, L. A. C.; EZAZ, T.;
14 JEGEDE, O. O. I.; LIEHR, T.; OLALEYE, V. F.; CIOFFI, M. B. First chromosomal analysis
15 in *Gymnarchus niloticus* (Gymnarchidae: Osteoglossiformes): insights into the karyotype
16 evolution of this ancient fish order. **Biological Journal of the Linnean Society**, v. 125,
17 p. 83–92, 2018. <https://doi.org/10.1093/biolinnean/bly098>
- 18 HÉNAFF, E.; VIVES, C.; DESVOYES, B.; CHAURASIA, A.; PAYET, J.; GUTIERREZ, C.;
19 CASACUBERTA, J. Extensive amplification of the E2F transcription factor binding sites
20 by transposons during evolution of *Brassica* species. **Pant Journal**, v. 77, p. 852–862,
21 2014. <https://doi.org/10.1111/tpj.12434>
- 22 HERNANDEZ, N. Small nuclear RNA genes: a model system to study fundamental
23 mechanisms of transcription. **J Biol Chem**, v. 276, p. 26733–26736, 2001.
24 <https://doi.org/10.1074/jbc.R100032200>
- 25 HICKMAN, A. B.; DYDA, F. DNA transposition at work. **Chem Rev**, v. 116, p. 12758–
26 12784, 2016. <https://doi.org/10.1021/acs.chemrev.6b00003>
- 27 HU, B.; ZHONG, L.; WENG, Y.; PENG, L.; HUANG, Y.; ZHAO, Y.; LIANG, X. Therapeutic
28 siRNA: state of the art. **Signal Transduct Target Ther**, v. 5, n. 101, 2020.
29 <https://doi.org/10.1038/s41392-020-0207-x>
- 30 INGENITO, L. F. S. **Análise filogenética da família Parodontidae (Teleostei,**
31 **Characiformes)**. 2008, 127f. Tese de Doutorado (Programa de Pós-graduação em
32 Zoologia). Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- 33 IZSVÁK, Z.; IVICS, Z.; SHIMODA, N.; MOHN, D.; OKAMOTO, H.; HACKETT, P. B. Short
34 inverted-repeat transposable elements in the teleost fish and implications for a
35 mechanism of their amplification. **J Mol Evol**, v. 48, p. 13–21, 1999.
36 <https://doi.org/10.1007/pl00006440>

- 1 JAWDEKAR, G. W.; HENRY, R. W. Transcriptional regulation of human small nuclear
2 RNA genes. **Biochim Biophys Acta**, v. 1779, p. 295–305, 2008.
3 <https://doi.org/10.1016/j.bbagr.2008.04.001>
- 4 JECK, W. R.; SORRENTINO, J. A.; WANG, K.; SLEVIN, M. K.; BURD, C. E.; LIU, J.;
5 MARZLUFF, W. F.; SHARPLESS, N. E. Circular RNAs are abundant, conserved, and
6 associated with ALU repeats. **RNA**, v. 19, p. 141–157, 2013.
7 <https://doi.org/10.1261/rna.035667.112>
- 8 JESUS, C. M.; BERTOLLO, L. A. C.; MOREIRA-FILHO, O. Comparative cytogenetics in
9 *Apareiodon affinis* (Pisces, Characiformes) and considerations regarding diversification of
10 the group. **Genetica**, v. 105, p. 63–67, 1999. <https://doi.org/10.1023/A:1003592022927>
- 11 JESUS, C. M.; MOREIRA-FILHO, O. Cytogenetic studies in some *Apareiodon* species
12 (Pisces, Parodontidae). **Cytologia**, v. 65, 4, p. 397–402, 2000a.
13 <https://doi.org/10.1508/cytologia.65.398>
- 14 JESUS, C. M.; MOREIRA-FILHO, O. Karyotypes of three species of *Parodon* (Teleostei,
15 Parodontidae). **Ichthyological Exploration Freshwaters**, v. 11, p. 75-80, 2000b.
- 16 JING, N.; BAO, Z.; ZHANG, X.; EDDY, S. R.; WESSLER, S. R. Pack-MULE transposable
17 elements mediate gene evolution in plants. **Nature**, v. 431, p. 569–573, 2004.
18 <https://doi.org/10.1038/nature02953>
- 19 JORGE, L. C.; MOREIRA-FILHO, O. Nucleolar organizer regions as markers of
20 chromosomal polymorphism in *Apareiodon affinis* (Pisces, Parodontidae). **Caryologia**, v.
21 57, p. 203–207, 2004. <https://doi.org/10.1080/00087114.2004.10589392>
- 22 KALVARI, I.; NAWROCKI, E. P.; ARGASINSKA, J.; QUINONES-OLIVEIRA, N.; FINN, R.
23 D.; BATEMAN, A.; PETROV, A. I. Non-coding RNA analysis using the Rfam database.
24 **Curr Protoc Bioinformatics**, v. 62, n. e51, 2018. <https://doi.org/10.1002/cpbi.51>
- 25 KALVARI, I.; NAWROCKI, E. P.; ONTIVEROS-PALACIOS, N.; ARGASINSKA, J.;
26 LAMKIEWICZ, K.; MARZ, M.; GRIFFITHS-JONES, S.; TOFFANO-NIOCHE, C.;
27 GAUTHERET, D.; WEINBERG, Z.; RIVAS, E.; EDDY, S. R.; FINN, R. D.; BATEMAN, A.;
28 PETROV, A. I. Rfam 14: expanded coverage of metagenomic, viral and microRNA
29 families. **Nucleic Acids Res**, v. 49, p. D192–D200, 2020.
30 <https://doi.org/10.1093/nar/gkaa1047>
- 31 KEARSE, M.; MOIR, R.; WILSON, A.; STONES-HAVAS, S.; CHEUNG, M.; STURROCK,
32 S.; BUXTON, S.; COOPER, A.; MARKOWITZ, S.; DURAN, C.; THIERER, T.; ASHTON,
33 B.; MEINTJES, P.; DRUMMOND, A. Geneious Basic: an integrated and extendable
34 desktop software platform for the organization and analysis of sequence data.
35 **Bioinformatics**, v. 28, p. 1647–1649, 2012. <https://doi.org/10.1093/bioinformatics/bts199>

- 1 KEHRER-SAWATZKI, H.; COOPER, D. N. Molecular mechanisms of chromosomal
2 rearrangement during primate evolution. **Chromosome Res**, v. 16, p. 41–56, 2008.
3 <https://doi.org/10.1007/s10577-007-1207-1>
- 4 KIDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes.
5 **Genetica**, v. 115, p. 49–63, 2002. <https://doi.org/10.1023/A:1016072014259>
- 6 KIDWELL, M. G.; LISCH, D. Transposable elements as sources of variation in animals
7 and plants. **PNAS**, v. 94, p. 7704–7711, 1997. <https://doi.org/10.1073/pnas.94.15.7704>
- 8 KLAI, K.; CHÉNAIS, B.; ZIDI, M.; DJEBBI, S.; CARUSO, A.; DENIS, F.; CONFAIS, J.;
9 BADAWI, M.; CASSE, N.; KHEMAKHEM, M. Screening of *Helicoverpa armigera*
10 mobilome revealed transposable element insertions in insecticide resistance genes.
11 **Insects**, v. 11, n. 879, 2020. <https://doi.org/10.3390/insects11120879>
- 12 KLAI, K.; ZIDI, M.; CHÉNAIS, B.; DENIS, F.; CARUSO, A.; CASSE, N.; KHEMAKHEM,
13 M. M. Miniature inverted-repeat transposable elements (MITEs) in the two Lepidopteran
14 genomes of *Helicoverpa armigera* and *Helicoverpa zea*. **Insects**, v. 13, n. 313, 2022.
15 <https://doi.org/10.3390/insects13040313>
- 16 KOHANY, O.; GENTLES, A. J.; HANKUS, L.; JURKA, J. Annotation, submission and
17 screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **BMC**
18 **Bioinformatics**, v. 7, n. 474, 2006. <https://doi.org/10.1186/1471-2105-7-474>
- 19 KOJIMA, K. K. A new class of SINEs with snRNA gene-derived-heads. **Genome Biol**
20 **Evol**, v. 7, p. 1702–1712, 2015. <https://doi.org/10.1093/gbe/evv100>
- 21 KOJIMA, K. K. Structural and sequence diversity of eukaryotic transposable elements.
22 **Genes & Genetic Systems**, v. 94, p. 233–252, 2019.
- 23 KOJIMA, K. K.; JURKA, J. A superfamily of DNA transposons targeting multicopy small
24 RNA genes. **PLoS One**, v. 8, n. e68260, 2013.
25 <https://doi.org/10.1371/journal.pone.0068260>
- 26 KONING, A. P. J.; GU, W.; CASTOE, T. A.; BATZER, M. A.; POLLOCK, D. D. Repetitive
27 elements may comprise over two-thirds of the human genome. **PLoS ONE**, v. 7, n.
28 e1002384, 2011. <https://doi.org/10.1371/journal.pgen.1002384>
- 29 KOREN, S.; WALENZ, B. P.; BERLIN, K.; MILLER, J. R.; BERGMAN, N. H.; PHILLIPPY,
30 A. M. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and
31 repeat separation. **Genome Res**, v. 27, p. 722–736, 2017.
32 <https://doi.org/10.1101/gr.215087.116>
- 33 KOREN, S.; RHIE, A.; WALENZ, B. P.; DILTHEY, A. T.; BICKHART, D. M.; KINGAN, S.
34 B.; HIENDLEDER, S.; WILLIAMS, T. P. L.; PHILLIPPY, A. M. *De novo* assembly of

- 1 haplotype-resolved genomes with trio binning. **Nat Biotechnol**, v. 36, p. 1174–1182,
2 2018. <https://doi.org/10.1038/nbt.4277>
- 3 KORNIENKO, A. E.; GUENZL, P. M.; BARLOW, D. P.; PAULER, F. M. Gene regulation
4 by the act of long non-coding RNA transcription. **BMC Biol**, v. 11, n. 59, 2013.
5 <https://doi.org/10.1186/1741-7007-11-59>
- 6 KOVALENKO, T. F.; PATRUSHEV, L. I. Pseudogenes as functionally significant elements
7 of the genome. **Biochemistry**, v. 83, p. 1332–1349, 2018.
8 <https://doi.org/10.1134/S0006297918110044>
- 9 KOZOMARA, A.; BIRGAOANU, M.; GRIFFITHS-JONES, S. miRBase: from microRNAs
10 sequences to function. **Nucleic Acids Res**, v. 47, p. D155–D162, 2019.
11 <https://doi.org/10.1093/nar/gky1141>
- 12 KUANG, H.; PADMANABHAN, C.; LI, F.; KAMEI, A.; BHASKAR, P. B.; OUYANG, S.;
13 JIANG, J.; BUELL, C. R.; BAKER, B. Identification of miniature inverted-repeat
14 transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: New
15 functional implications for MITEs. **Genome Res**, v. 19, p. 42–56, 2009. <https://doi.org/10.1101/gr.078196.108>
- 17 LERNER, M. R.; BOYLE, J. A.; MOUNT, S. M.; WOLIN, S. L.; STEITZ, J. A. Are snRNPs
18 involved in splicing? **Nature**, v. 283, p. 220–224, 1980. <https://doi.org/10.1038/283220a0>
- 19 LEVAN, A.; FREDGA, K.; SANDBERG, A. A. Nomenclature for centromeric position on
20 chromosomes. **Hereditas** 52:201–220, 1964. <https://doi.org/10.1111/j.1601-5223.1964.tb01953.x>
- 22 LI, W.; YANG, W.; WANG, X. J. Pseudogenes: pseudo or real functional elements? **J**
23 **Genet Genomics**, v. 40, p. 171–177, 2013. <https://doi.org/10.1016/j.jgg.2013.03.003>
- 24 LIU, G.; MATTICK, J. S.; TAFT, R. J. A meta-analysis of the genomic and transcriptomic
25 composition of complex life. **Cell Cycle**, v. 12, p. 2061–2072, 2013.
26 <https://doi.org/10.4161/cc.25134>
- 27 LIU, Y.; QAMAR, M. T.; FENG, J.; DING, Y.; WANG, S.; WU, G.; KE, L.; XU, Q.; CHEN,
28 L. Comparative analysis of miniature inverted-repeat transposable elements (MITEs) and
29 long terminal repeat (LTR) retrotransposons in six *Citrus* species. **BMC Plant Biology**, v.
30 19, n. 140, 2019. <https://doi.org/10.1186/s12870-019-1757-3>
- 31 LÓPEZ, M. D.; ROSENBLAD, M. A.; SAMUELSSON, T. Computational screen for
32 spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor
33 spliceosomal components. **Nucleic Acids Res**, v. 36, p. 3001–3010, 2008.
34 <https://doi.org/10.1093/nar/gkn142>

- 1 LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. A. The repetitive DNA content of Eukaryotic
2 Genomes. **Genome Dyn**, v. 7, p. 1–28, 2012. <https://doi.org/10.1159/000337118>
- 3 LORENZ, R.; BERNHART, S. H.; HÖNER, Z. U.; SIEDERDISSEN, C.; TAFER, H.;
4 FLAMM, C.; STADLER, P. F.; HOFACKER, I. L. ViennaRNA Package 2.0. **Algorithms**
5 **Mol Biol**, v. 6, n. 26, 2011. <https://doi.org/10.1186/1748-7188-6-26>
- 6 LORKOVIĆ, Z. J.; LEHNER, R.; FORSTNER, C.; BARTA, A. Evolutionary conservation
7 of the minor U12-type spliceosome between plants and humans. **RNA**, v. 11, p. 1095–
8 1107, 2005. <https://doi.org/10.1261/rna.2440305>
- 9 LU, C.; CHEN, J.; ZHANG, Y.; HU, Q.; SU, W.; KUANG, H. Miniature inverted-repeat
10 transposable elements (MITEs) have been accumulated through amplification bursts and
11 play important roles in gene expression and species diversity in *Oryza sativa*. **Mol Biol**
12 **Evol**, v. 29, p. 1005–1017, 2012. <https://doi.org/10.1093/molbev/msr282>
- 13 LU, T. X.; ROTHENBERG, M. E. MicroRNA. **Journal of Allergy and Clinical**
14 **Immunology**, v. 141, p. 1202–1207, 2018. <https://doi.org/10.1016/j.jaci.2017.08.034>
- 15 MACKO-PODGÓRNI, A.; MACHAJ, G.; GRZEBELUS, D. A global landscape of miniature
16 inverted-repeat transposable elements in the carrot genome. **Genes**, v. 12, n. 859, 2021.
17 <https://doi.org/10.3390/genes12060859>
- 18 MALIMPENSA, G. C.; TRALDI, J. B.; TOYAMA, D.; HENRIQUE-SILVA, F.; VICARI, M.
19 R.; MOREIRA-FILHO, O. Chromosomal Mapping of Repeat DNA in *Bergiarja*
20 *westermanni* (Pimelodidae, Siluriformes): Localization of 45S rDNA in B Chromosomes.
21 **Cytogenet Genome Res**, v. 154, p. 99–106, 2018. <https://doi.org/10.1159/000487652>
- 22 MALIMPENSA, G. C.; TRALDI, J. B.; MARTINEZ, J. F.; DEON, G.; AZAMBUJA, M.;
23 NOGAROTO, V.; VICARI, M. R.; MOREIRA-FILHO, O. Chromosomal Diversification in
24 Two Species of *Pimelodus* (Siluriformes: Pimelodidae): Comparative Cytogenetic
25 Mapping of Multigene Families. **Zebrafish**, v. 17, p. 278–286, 2020.
26 <https://doi.org/10.1089/zeb.2020.1892>
- 27 MANCHADO, M.; ZUASTI, E.; CROSS, I.; MERLO, A.; INFANTE, C.; REBORDINOS, L.
28 Molecular characterization and chromosomal mapping of the 5S rRNA gene in *Solea*
29 *senegalensis*: a new linkage to the U1, U2, and U5 small nuclear genes. **Genome**, v. 49,
30 p. 79–86, 2006. <https://doi.org/10.1139/g05-068>
- 31 MANEECHOT, N.; YANO, C. F.; BERTOLLO, L. A. C.; GETLEKHA, N.; MOLINA, W. F.;
32 DITCHAROEN, S.; TENGJAROENKUL, B.; SUPIWONG, W.; TANOMTONG, A.; CIOFFI,
33 M. B. Genomic organization of repetitive DNAs highlights chromosomal evolution in the
34 genus *Clarias* (Clariidae, Siluriformes). **Mol Cytogenet**, v. 9, n. 4, 2016.
35 <https://doi.org/10.1186/s13039-016-0215-2>

- 1 MARTINS, C.; CABRAL-DE-MELLO, D. C.; VALENTE, G. T.; MAZZUCHELLI, J.;
2 OLIVEIRA, S. G. Cytogenetic Mapping and Contribution to the Knowledge of Animal
3 Genomes. *In*: URBANO, K. V. (ed) *Advances in Genetics Research*, vol 4. Nova **Science**
4 **Publisher**, New York, pp 1–82, 2010.
- 5 MARZ, M.; KIRSTEN, T.; STADLER, P. F. Evolution of spliceosomal snRNA genes in
6 metazoan animals. **J Mol Evol**, v. 67, p. 594–607, 2008. [https://doi.org/10.1007/s00239-](https://doi.org/10.1007/s00239-008-9149-6)
7 008-9149-6
- 8 MCCLINTOCK, B. The origin and behavior of mutable loci in maize. **Biological sciences**,
9 v. 36, p. 344–355, 1950.
- 10 MERLO, M. A.; CROSS, I.; CHAIRI, H.; MANCHADO, M.; REBORDINOS, L. Analysis of
11 three multigene families as useful tools in species characterization of two closely-related
12 species, *Dicentrarchus labrax*, *Dicentrarchus punctatus* and their hybrids. **Genes Genet**
13 **Syst**, v. 85, p. 341–349, 2010. <https://doi.org/10.1266/ggs.85.341>
- 14 MERLO, M. A.; PACCHIARINI, T.; PORTELA-BENS, S.; CROSS, I.; MANCHADO, M.;
15 REBORDINOS, L. Genetic characterization of *Plectorhinchus mediterraneus* yields
16 important clues about genome organization and evolution of multigene families. **BMC**
17 **Genet**, v. 13, n. 33, 2012a. <https://doi.org/10.1186/1471-2156-13-33>
- 18 MERLO, M. A.; CROSS, I.; RODRÍGUEZ-RÚA, A.; MANCHADO, M.; REBORDINOS, L.
19 First approach to studying the genetics of the meagre (*Argyrosomus regius*; Asso, 1801)
20 using three multigene families. **Aquaculture Research**, v. 44, p. 974–984, 2012b.
21 <https://doi.org/10.1111/j.1365-2109.2012.03103.x>
- 22 MERLO, M. A.; CROSS, I.; MANCHADO, M.; CÁRDENAS, S.; REBORDINOS, L. The 5S
23 rDNA high dynamism in *Diplodus sargus* is a Transposon-mediated mechanism.
24 Comparison with other multigene families and Sparidae species. **J Mol Evol**, v. 76, p. 83–
25 97, 2013. <https://doi.org/10.1007/s00239-013-9541-8>
- 26 MEWES, H. W.; ALBERMANN, K.; BÄHR, M.; FRISHMAN, D.; GLEISSNER, A.; HANI,
27 J.; HEUMANN, K.; KLEINE, K.; MAIERL, A.; OLIVER, S. G.; PFEIFFER, F.; ZOLLNER,
28 A. Overview of the yeast genome. **Nature**, v. 387, p. 7–8, 1997.
29 <https://doi.org/10.1038/387s007>
- 30 MIGHELL, A. J.; SMITH, N. R.; ROBINSON, P. A.; MARKHAM, A. F. Vertebrate
31 pseudogenes. **FEBS Lett**, v. 468, p. 109–114, 2000. [https://doi.org/10.1016/S0014-](https://doi.org/10.1016/S0014-5793(00)01199-6)
32 5793(00)01199-6
- 33 MIKKELSEN, E.; WEIR, J. T. The genome of the Xingu scale-backed antbird (*Willisornis*
34 *vidua nigrigula*) reveals lineage-specific adaptations. **Genomics**, v. 112, p. 4552–4560,
35 2020. <https://doi.org/10.1016/j.ygeno.2020.07.047>

- 1 MILHOMEM, S. S. R.; SCACCHETTI, P. C.; PIECZARKA, J. C.; FERGUSON-SMITH, M.
2 A.; PANSONATO-ALVES, J. C.; O'BRIEN, P. C. M.; FORESTI, F.; NAGAMACHI, C. Y.
3 Are NORs always located on homeologous chromosomes? A FISH investigation with
4 rDNA and whole chromosome probes in *Gymnotus* fishes (Gymnotiformes). **PLoS One**,
5 v. 8, n. e55608, 2013. <https://doi.org/10.1371/journal.pone.0055608>
- 6 MILLER, W. J.; MCDONALD, J. F.; PINSKER, W. Molecular domestication of mobile
7 elements. *In*: CAPY, P. (Ed). Evolution and impact of transposable elements.
8 **Contemporary Issues in Genetics and evolution**, v. 6, Dordrecht: Springer, p. 261–
9 270, 1997.
- 10 MORATA, J.; MARÍN, F.; PAYET, J.; CASACUBERTA, J. M. Plant lineage-specific
11 amplification of transcription factor binding motifs by miniature inverted-repeat
12 transposable elements (MITEs). **Genome Biology Evol**, v. 10, p. 1210–1220, 2018.
13 <https://doi.org/10.1093/gbe/evy073>
- 14 MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; GALETTI Jr, P. M. Evidences for a Multiple
15 Sex Chromosome System with Female Heterogamety in *Apareiodon Affinis* (Pisces,
16 Parodontidae). **Caryologia**, v. 33, p. 83–91, 1980.
17 <https://doi.org/10.1080/00087114.1980.10796821>
- 18 MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; GALETTI Jr, P. M. Distribution of sex
19 chromosome mechanism in neotropical fish and description of a ZZ/ZW system in
20 *Parodon hilarii* (Parodontidae). **Caryologia**, v. 46, p. 115–125, 1993.
21 <https://doi.org/10.1080/00087114.1993.10797253>
- 22 MURRAY, M. G.; THOMPSON, W. F. Rapid isolation of high molecular weight plant DNA.
23 **Nucleic Acids Res**, v. 8, p. 4321–4326, 1980. <https://doi.org/10.1093/nar/8.19.4321>
- 24 NAKATANI, K.; AGOSTINHO, A. A.; BAUMGARTNER, G.; BIALETZKI, A.; SANCHES,
25 P. V.; MAKRAKIS, M.C.; PAVANELLI, C. S. **Ovos e larvas de água doce:
26 desenvolvimento e manual de identificação**. Maringá: EDUEM, p.378, 2001.
- 27 NASCIMENTO, V. D.; COELHO, K. A.; NOGAROTO, V.; ALMEIDA, R. B.; ZIEMNICZAK,
28 K.; CENTOFANTE, L.; PAVANELLI, C. S.; TORRES, R. A.; MOREIRA-FILHO, O.;
29 VICARI, M. R. Do multiple karyomorphs and population genetics of freshwater darter
30 characines (*Apareiodon affinis*) indicate chromosomal speciation? **Zool Anz**, v. 272, p.
31 93–103, 2018. <https://doi.org/10.1016/j.jcz.2017.12.006>
- 32 NAWROCKI, E. P.; EDDY, S. R. Infernal 1.1: 100-fold faster RNA homology searches.
33 **Bioinformatics**, v. 29, p. 2933–2935, 2013. <https://doi.org/10.1093/bioinformatics/btt509>
- 34 NEI, M.; ROONEY, A. P. Concerted and birth-and-death evolution of multigene families.
35 **Annu Rev Genet**, v. 39, p. 121–152, 2005.
36 <https://doi.org/10.1146/annurev.genet.39.073003.112240>

- 1 NIRCHIO, M.; MASACHE, M. C.; PAIM, F. G.; CIOFFI, M. B.; MOREIRA-FILHO, O.;
2 BARRIGA, R.; OLIVEIRA, C.; ROSSI, A. R. Chromosome analysis in *Saccodon wagneri*
3 (Characiformes) and insights into the karyotype evolution of Parodontidae. **Neotrop**
4 **Ichthyol**, v. 19, n. e200103, 2021. <https://doi.org/10.1590/1982-0224-2020-0103>
- 5 NISHIMURA, O.; HARA, Y.; KURAKU, S. gVolante for standardizing completeness
6 assessment of genome and transcriptome assemblies. **Bioinformatics**, v. 33, p. 3635–
7 3637, 2017. <https://doi.org/10.1093/bioinformatics/btx445>
- 8 NOTO, J. J.; SCHMIDT, C. A.; MATERA, A. G. Engineering and expressing circular RNAs
9 via tRNA splicing. **RNA Biol**, v. 14, p. 978–984, 2017.
10 <https://doi.org/10.1080/15476286.2017.1317911>
- 11 NURK, S.; WALENZ, B. P.; RHIE, A.; VOLLGER, M. R.; LOGSDON, G. A.; GROTHE, R.;
12 MIGA, K. H.; EICHLER, E. E.; PHILLIPPY, A. M.; KOREN, S. HiCanu: accurate assembly
13 of segmental duplications, satellites, and allelic variants from high-fidelity long reads.
14 **Genome Res**, v. 30, p. 1291–1305, 2020. <https://doi.org/10.1101/gr.263566.120>
- 15 OLIVEIRA, E. A.; SASSI, F. M. C.; PEREZ, M. F.; BERTOLLO, L. A. C.; RÁB, P.; EZAZ,
16 T.; HATANAKA, T.; VIANA, P. F.; FELDBERG, E.; OLIVEIRA, E. H. C.; CIOFFI, M. B.
17 Comparative cytogenetic survey of the giant bonytongue *Arapaima* fish
18 (Osteoglossiformes: Arapaimidae), across different Amazonian and Tocantins/Araguaia
19 River basins. **Neotrop Ichthyol**, v. 18, n. e200055, 2020. <https://doi.org/10.1590/1982-0224-2020-0055>
- 21 PAIM, F. G.; OLIVEIRA NOBILE, M. L. M.; FORESTI, F.; OLIVEIRA, C. Cytogenetic Tools
22 to Study the Biodiversity of Neotropical Fish: From the Classic to the Advent of Cell
23 Culture. *In*: LARRAMENDY, M.; SOLONESKI, S. (ed) Cytogenetics-Past, Present and
24 Further Perspectives. **IntechOpen**, London, p. 53–73, 2018
- 25 PATEL, A.; STEITZ, J. Splicing double: insights from the second spliceosome. **Nat Rev**
26 **Mol Cell Biol**, v. 4, p. 960–970, 2003. <https://doi.org/10.1038/nrm1259>
- 27 PAUL, S. C.; SHARMA, A.; MEHTA, R.; PAUL, S. Genome wide computational
28 identification of tuna (*Thunnus orientalis*) microRNAs and their targets. **Ocean Sci J**, v.
29 53, p. 727–734, 2018. <https://doi.org/10.1007/s12601-018-0041-z>
- 30 PAULA, A. A.; PENHA, H. A.; DELAI, V. A.; GIULIANO-CAETANO, L.; DIAS, A. L.
31 Occurrence of structural polymorphism and supernumerary chromosomes in a population
32 of *Parodon nasus* (Parodontidae). **Caryologia**, v. 70, p. 200–205, 2017.
33 <https://doi.org/10.1080/00087114.2017.1318503>
- 34 PAVANELLI, C. S. Family Parodontidae. *In*: REIS, R. E.; KULLANDER, S. O.; FERRARIS,
35 C. J. (Eds.). Check list of the freshwater fishes of South and Central America. Porto
36 Alegre: **EDIPUCRS**, p. 46–50, 2003

- 1 PAVANELLI, C. S.; BRITSKI, H. A. *Apareiodon* Eigenmann, 1916 (Teleostei,
2 Characiformes), from the Tocantins-Araguaia Basin, with description of three new
3 species. **Copeia**, v. 2003, p. 337–348, 2003. [https://doi.org/10.1643/0045-](https://doi.org/10.1643/0045-8511(2003)003[0337:AETCFT]2.0.CO;2)
4 [8511\(2003\)003\[0337:AETCFT\]2.0.CO;2](https://doi.org/10.1643/0045-8511(2003)003[0337:AETCFT]2.0.CO;2)
- 5 PESCHANSKY, V. J.; WAHLESTEDT, C. Non-coding RNAs as direct and indirect
6 modulators of epigenetic regulation. **Epigenetics**, v. 9, p. 3–12. 2014.
7 <https://doi.org/10.4161/epi.27473>
- 8 PINKEL, D.; STRAUME, T.; GRAY, J. W. Cytogenetic analysis using quantitative, high-
9 sensitivity, fluorescence hybridization. **Proc Natl Acad Sci U S A**, v. 83, p. 2934–2938,
10 1986. <https://doi.org/10.1073/pnas.83.9.2934>
- 11 PIRIYAPONGSA, J.; JORDAN, K. Dual coding of siRNAs and miRNAs by plant
12 transposable elements. **RNA**, v. 14, p. 814–821, 2008. [https://doi.org/](https://doi.org/10.1261/rna.916708)
13 [10.1261/rna.916708](https://doi.org/10.1261/rna.916708)
- 14 PISCOR, D.; CENTOFANTE, L.; PARISE-MALTEMPI, P. P. Highly similar morphologies
15 between chromosomes bearing U2 snRNA gene clusters in the group *Astyanax* Baird and
16 Girard, 1854 (Characiformes, Characidae): An evolutionary approach in species with 2n
17 = 36, 46, 48 and 50. **Zebrafish**, v. 13, p. 565–570, 2016.
18 <https://doi.org/10.1089/zeb.2016.1292>
- 19 PISCOR, D.; FERNANDES, C. A.; PARISE-MALTEMPI, P. P. Conserved number of U2
20 snDNA sites in *Piabina argentea*, *Piabarchus stramineus* and two *Bryconamericus*
21 species (Characidae, Stevardiinae). *Neotrop Ichthyol* v. 16; n. e170066, 2018.
22 <https://doi.org/10.1590/1982-0224-20170066>
- 23 PONJAVIC, J.; PONTING, C. P.; LUNTER, G. Functionality or transcriptional noise?
24 Evidence for selection within long noncoding RNAs. **Genome Res**, v. 17, p. 556–565,
25 2007. <https://doi.org/10.1101/gr.6036807>
- 26 PONZIO, J. C.; PISCOR, D.; PARISE-MALTEMPI, P. P. Chromosomal locations of U2
27 snDNA clusters in *Megaleporinus*, *Leporinus* and *Schizodon* (Characiformes:
28 Anostomidae). **Biologia**, v. 73, p. 295–298, 2018. [https://doi.org/10.2478/s11756-018-](https://doi.org/10.2478/s11756-018-0031-8)
29 [0031-8](https://doi.org/10.2478/s11756-018-0031-8)
- 30 POUTER, R. T. M.; BUTLER, M. I. Tyrosine recombinase retrotransposons and
31 transposon. In CRAIG, N.; CHANDLER, M.; GELLERT, M.; LAMBOWITZ, A.; RICE, P.;
32 SANDMEYER, S. (Eds) Mobile DNA III. **American Society of Microbiology**:
33 Washington, p. 1271–1291, 2015.
- 34 PRYSZCZ, L. P.; GABALDÓN, T. Redundans: an assembly pipeline for highly
35 heterozygous genomes. **Nucleic Acids Res**, v. 44, p. e113, 2016.
36 <https://doi.org/10.1093/nar/gkw294>

- 1 PUCCI, M. B.; NOGAROTO, V.; MOREIRA-FILHO, O.; VICARI, M. R. Dispersion of
2 transposable elements and multigene families: Microstructural variation in *Characidium*
3 (Characiformes: Crenuchidae) genomes. **Genet Mol Biol**, v. 41, p. 585–592, 2018.
4 <https://doi.org/10.1590/1678-4685-GMB-2017-0121>
- 5 QUINLAN, A. R.; HALL, I. M. BEDTools: a flexible suite of utilities for comparing genomic
6 features. **Bioinformatics**, v. 26, p. 841–842, 2010.
7 <https://doi.org/10.1093/bioinformatics/btq033>
- 8 RÁB, P.; YANO, C. F.; LAVOUÉ, S.; JEGEDE, O. I.; BERTOLLO, L. A. C.; EZAZ, T.;
9 MAJTÁNOVÁ, Z.; OLIVEIRA, E. A.; CIOFFI, M. B. Karyotype and mapping of repetitive
10 DNAs in the African butterfly fish *Pantodon buchholzi*, the sole species of the family
11 Pantodontidae. **Cytogenet Genome Res**, v. 149, p. 312–320, 2016.
12 <https://doi.org/10.1159/000450534>
- 13 ROGNES, T.; FLOURI, T.; NICHOLS, B.; QUINCE, C.; MAHÉ, F. VSEARCH: a versatile
14 open source tools for metagenomics. **PeerJ**, v. 4, n. e2584, 2016.
15 <https://doi.org/10.7717/peerj.2584>
- 16 ROSA, R.; BELLAFRONTÉ, E.; MOREIRA-FILHO, O.; MARGARIDO, V. P. Constitutive
17 heterochromatin, 5S and 18S rDNA genes in *Apareiodon* sp. (Characiformes,
18 Parodontidae) with a ZZ/ZW sex chromosomes system. **Genetica**, v. 128, p. 159–166,
19 2006. <https://doi.org/10.1007/s10709-005-5700-1>
- 20 RÖTHER, S.; MEISTER, G. Small RNAs derived from longer non-coding RNAs.
21 **Biochimie**, v. 93, p. 1905–1915, 2011. <https://doi.org/10.1016/j.biochi.2011.07.032>
- 22 RUSSEL, A. G.; CHARETTE, J. M.; SPENCER, D. F.; GRAY, M. W. An early evolutionary
23 origin for the minor spliceosome. **Nature**, v. 443, p. 863–866, 2006.
24 <https://doi.org/10.1038/nature05228>
- 25 SAETROM, P.; SNOVE, O.; NEDLAND, M.; GRÜNFELD, T. B.; LIN, Y.; BASS, M. B.
26 CANON, J. R. Conserved microRNA characteristics in mammals. **Oligonucleotides**, v.
27 16, p. 115 – 144, 2006. <https://doi.org/10.1089/oli.2006.16.115>
- 28 SANTOS, E. O.; DEON, G. A.; ALMEIDA, R. B.; OLIVEIRA, E. A.; NOGAROTO, V.;
29 SILVA, H. P.; PAVANELLI, C. S.; CESTARI, M. M.; BERTOLLO, L. A. C.; MOREIRA-
30 FILHO, O.; VICARI, M. R. Cytogenetics and DNA barcode reveal an undescribed
31 *Apareiodon* species (Characiformes: Parodontidae). **Genet Mol Biol**, v. 42, p. 365–373,
32 2019. <https://doi.org/10.1590/1678-4685-GMB-2018-0066>
- 33 SCACCHETTI, P. C.; UTSUNOMIA, R.; PANSONATO-ALVES, J. C.; COSTA SILVA, G.
34 J.; VICARI, M. R.; ARTONI, R. F.; OLIVEIRA, C.; FORESTI, F. Repetitive DNA sequences
35 and evolution of ZZ/ZW sex chromosomes in *Characidium* (Teleostei: Characiformes).
36 **PLoS One**, v. 10, n. e0137231, 2015. <https://doi.org/10.1371/journal.pone.0137231>

- 1 SCHEMBERGER, M. O.; BELLAFRONTE, E.; NOGAROTO, V.; ALMEIDA, M. C.;
2 SCHÜHLI, G. S.; ARTONI, R. F.; MOREIRA-FILHO, O.; VICARI, M. R. Differentiation of
3 repetitive DNA sites and sex chromosome systems reveal closely related group in
4 Parodontidae (Actinopterygii: Characiformes). **Genetica**, v. 139, p. 1499–1508, 2011.
5 <https://doi.org/10.1007/s10709-012-9649-6>
- 6 SCHEMBERGER, M. O.; OLIVEIRA, J. I. N.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI,
7 R. F.; CESTARI, M. M.; MOREIRA-FILHO, O.; VICARI, M. R. Construction and
8 characterization of a repetitive DNA library in Parodontidae (Actinopterygii:
9 Characiformes): A genomic and evolutionary approach to the degeneration of the W sex
10 chromosome. **Zebrafish**, v. 11, p. 518–527, 2014. <https://doi.org/10.1089/zeb.2014.1013>
- 11 SCHEMBERGER, M. O.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI, R. F.; VALENTE,
12 G. T.; MARTINS, C.; MOREIRA-FILHO, O.; CESTARI, M. M.; VICARI, M. R. Sequence
13 analyses and chromosomal distribution of the *Tc1/Mariner* element in Parodontidae fish
14 (Teleostei: Characiformes). **Gene**, v. 593, p. 308–314, 2016.
15 <https://doi.org/10.1016/j.gene.2016.08.034>
- 16 SCHEMBERGER, M. O.; NASCIMENTO, V. D.; COAN, R.; RAMOS, E.; NOGAROTO,
17 V.; ZIEMNICZAK, K.; VALENTE, G. T.; MOREIRA-FILHO, O.; MARTINS, C.; VICARI, M.
18 R. DNA transposon invasion and microsatellite accumulation guide W chromosome
19 differentiation in a Neotropical fish genome. **Chromosoma**, v. 128, p. 547–560, 2019.
20 <https://doi.org/10.1007/s00412-019-00721-9>
- 21 SEMBER, A.; BOHLEN, J.; ŠLECHTOVÁ, V.; ALTMANOVÁ, M.; PELIKÁNOVÁ, S.; RÁB,
22 P. Dynamics of tandemly repeated DNA sequences during evolution of diploid and
23 tetraploid botiid loaches (Teleostei: Cobitoidea: Botiidae). **PLoS One**, v. 13, n. e0195054,
24 2018. <https://doi.org/10.1371/journal.pone.0195054>
- 25 SERRANO, É. A.; UTSUNOMIA, R.; SCUDELLER, O. S.; OLIVEIRA, C.; FORESTI, F.
26 Origin of B chromosomes in *Characidium alipioi* (Characiformes, Crenuchidae) and its
27 relationship with supernumerary chromosomes in other *Characidium* species. **Comp**
28 **Cytogenet**, v. 11, p. 81–95, 2017. <https://doi.org/10.3897/CompCytogen.v11i1.10886>
- 29 SHETH, N.; ROCA, X.; HASTINGS, M. L.; ROEDER, T.; KRAINER, A. R.;
30 SACHIDANANDAM, R. Comprehensive splice-site analysis using comparative genomics.
31 **Nucleic Acids Res**, v. 34, p. 3955–3967, 2006. <https://doi.org/10.1093/nar/gkl556>
- 32 SHI, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome.
33 **Nat Rev Mol Cell Biol**, v. 18, p. 655–670, 2017. <https://doi.org/10.1038/nrm.2017.86>
- 34 SILVA, D. M. Z. A.; UTSUNOMIA, R.; PANSONATO-ALVES, J. C.; OLIVEIRA, C.;
35 FORESTI, F. Chromosomal mapping of repetitive DNA sequences in five species of
36 *Astyanax* (Characiformes, Characidae) reveals independent location of U1 and U2 snRNA
37 sites and association of U1 snRNA and 5S rDNA. **Cytogenet Genome Res**, v. 146, p.
38 144–152, 2015. <https://doi.org/10.1159/000438813>

- 1 SIMÃO, F. A.; WATERHOUSE, R. M.; IOANNIDIS, P.; KRIVENTSEVA, E. V.; ZDOBNOV,
2 E. M. BUSCO: assessing genome assembly and annotation completeness with single-
3 copy orthologs. **Bioinformatics**, v. 31, p. 3210–3212, 2015.
4 <https://doi.org/10.1093/bioinformatics/btv351>
- 5 SIOMI, M. C.; SATO, K.; PEZIC, D.; ARAVIN, A. A. PIWI-interacting small RNAs: the
6 vanguard of genome defence. **Nat Rev Mol Cell Biol**, v. 12, p. 246–258, 2011.
7 <https://doi.org/10.1038/nrm3089>
- 8 SOARES, L. B.; PAIM, F. G.; RAMOS, L. P.; FORESTI, F.; OLIVERIA, C. Molecular
9 cytogenetic analysis and the establishment of a cell culture in the fish species
10 *Hollandichthys multifasciatus* (Eigenmann & Norris, 1900) (Characiformes, Characidae).
11 **Genet Mol Biol**, v. 44, n. e20200260, 2021. [https://doi.org/10.1590/1678-4685-GMB-
12 2020-0260](https://doi.org/10.1590/1678-4685-GMB-2020-0260)
- 13 SOCHOROVÁ, J.; GARCIA, S.; GÁLVEZ, F.; SYMONOVÁ, R.; KOVAŘÍK, A. Evolutionary
14 trends in animal ribosomal DNA loci: introduction to a new online database.
15 **Chromosoma** v. 127, p. 141–150, 2018. <https://doi.org/10.1007/s00412-017-0651-8>
- 16 SOTERO-CAIO, C. G.; PLATT, R. N.; SUH, A.; RAY, D. A. Evolution and diversity of
17 transposable elements in vertebrate genomes. **Genome Biol Evol**, v. 9, p. 161–177,
18 2017. <https://doi.org/10.1093/gbe/evw264>
- 19 SOUZA e SOUSA, J. F.; VIANA, P. F.; BERTOLLO, L. A. C.; CIOFFI, M. B.; FELDBERG,
20 E. Evolutionary relationships among *Boulengerella* species (Ctenoluciidae,
21 Characiformes): Genomic organization of repetitive DNAs and highly conserved
22 karyotypes. **Cytogenet Genome Res**, v. 152, p. 194–203, 2017.
23 <https://doi.org/10.1159/000480141>
- 24 SUMNER, A. T. Chromosomes – organization and function. **Blackwell Publishing
25 Company**, London, 2003
- 26 SUPIWONG, W.; LIEHR, T.; CIOFFI, M. B.; CHAVEERACH, A.; KOSYAKOVA, N.;
27 PINTHONG, K.; TANEE, T.; TANOMTONG, A. Karyotype and cytogenetic mapping of 9
28 classes of repetitive DNAs in the genome of the naked catfish *Mystus bocourti*
29 (Siluriformes, Bagridae). **Mol Cytogenet** v. 6, p. 51, 2013. [https://doi.org/10.1186/1755-
30 8166-6-51](https://doi.org/10.1186/1755-8166-6-51)
- 31 TAKAGUI, F. H.; BAUMGÄRTNER, L.; VENTURELLI, N. B.; PAIZ, L. M.; VIANA, P.;
32 DIONÍSIO, J. F.; POMPEO, L. R. S.; MARGARIDO, V. P.; FENOCCHIO, A. S.; ROSA,
33 R.; GIULIANO-CAETANO, L. Unrevealing the karyotypic evolution and cytotaxonomy of
34 armored catfishes (Loricariinae) with emphasis in *Sturisoma*, *Loricariichthys*, *Loricaria*,
35 *Proloricaria*, *Pyxiloricaria*, and *Rineloricaria*. **Zebrafish**, v. 17, p. 319–332, 2020.
36 <https://doi.org/10.1089/zeb.2020.1893>

- 1 TAKAGUI, F. H.; BAUMGÄRTNER, L.; VIANA, P.; LIMA, M. C. C.; BITENCOURT, J. A.;
2 VENERE, P. C.; LUI, R. L.; MOREIRA-FILHO, O.; FELDBERG, E.; ALMEIDA SIMÕES,
3 F.; BIRINDELLI, J. L.; GIULIANO-CAETANO, L. Karyotype evolution of talking thorny
4 catfishes *Anadoras* (Doradidae, Astrodoradinae): a process mediated by structural
5 rearrangements and intense reorganization of repetitive DNAs. **Cytogenet Genome Res**,
6 v. 162, p. 64–75, 2022. <https://doi.org/10.1159/000523747>
- 7 TAKKI, O.; KOMISSAROV, A.; KULAK, M.; GALKINA, S. Identification of centromere-
8 specific repeats in the zebra finch genome. **Cytogenet Genome Res**, v. 162, p. 55–63,
9 2022. <https://doi.org/10.1159/000521716>
- 10 TESTORI, A.; CAIZZI, L.; CUTRUPI, S.; FRIARD, O.; BORTOLI, M.; CORA, D.;
11 CASELLE, M. The role of transposable elements in shaping the combinatorial interaction
12 of transcription factor. **BMC Genomics**, v. 13, n. 400, 2012. [https://doi.org/10.1186/1471-
13 2164-13-400](https://doi.org/10.1186/1471-2164-13-400)
- 14 TOLLIS, M.; BOISSINOT, S. The evolutionary dynamics of transposable elements in
15 eukaryote genomes. **Repetitive DNA**, v. 7, p. 68–91, 2012.
16 <https://doi.org/10.1159/000337126>
- 17 TRALDI, J. B.; VICARI, M. R.; MARTINEZ, J. F.; BLANCO, D. R.; LUI, R. L.; MOREIRA-
18 FILHO, O. Chromosome Analyses of *Apareiodon argenteus* and *Apareiodon davisii*
19 (Characiformes, Parodontidae): An Extensive Chromosomal Polymorphism of 45S and
20 5S Ribosomal DNAs. **Zebrafish**, v. 13, p. 19–25, 2016.
21 <https://doi.org/10.1089/zeb.2015.1124>
- 22 TRALDI, J. B.; ZIEMNICZAK, K.; FÁTIMA MARTINEZ, J.; BLANCO, D. R.; LUI, R. L.;
23 SCHEMBERGER, M. O.; NOGAROTO, V.; MOREIRA-FILHO, O.; VICARI, M. R.
24 Chromosome Mapping of H1 and H4 Histones in Parodontidae (Actinopterygii:
25 Characiformes): Dispersed and/or Co-Opted Transposable Elements? **Cytogenet
26 Genome Res**, v. 158, p. 106–113, 2019. <https://doi.org/10.1159/000500987>
- 27 TRALDI, J. B.; VICARI, M. R.; MARTINEZ, J. F.; BLANCO, D. R.; LUI, R. L.; AZAMBUJA,
28 M.; ALMEIDA, R. B.; MALIMPENSA, G. C.; COSTA SILVA, G. J.; OLIVEIRA, C.;
29 PAVANELLI, C. S.; MOREIRA-FILHO, O. Recent *Apareiodon* species evolutionary
30 divergence (Characiformes: Parodontidae) evidenced by chromosomal and molecular
31 inference. **Zool Anz**, v. 289, p. 166–176, 2020. <https://doi.org/10.1016/j.jcz.2020.10.010>
- 32 TURUNEN, J. J.; NIEMELÄ, E. H.; VERMA, B.; FRILANDER, M. J. The significant other:
33 splicing by the minor spliceosome. **Wiley Interdiscip Rev RNA**, v. 4, p. 61–76, 2013.
34 <https://doi.org/10.1002/wrna.1141>
- 35 ÚBEDA-MANZANARO, M.; MERLO, M. A.; PALAZÓN, J. L.; CROSS, I.; SARASQUETE,
36 C.; REBORDINOS, L. Chromosomal mapping of the major and minor ribosomal genes,
37 (GATA)_n and U2 snRNA gene by double-colour FISH in species of the Batrachoididae
38 family. **Genetica**, v. 138, p. 787–794, 2010. <https://doi.org/10.1007/s10709-010-9460-1>

- 1 UNTERGASSER, A.; NIJVEEN, H.; RAO, X.; BISSELING, T.; GEURTS, R.; LEUNISSEN,
2 J. A. M. Primer3Plus, an enhanced web interface to primer3. **Nucleic Acids Res**, v. 35,
3 p. W71–W74, 2007. <https://doi.org/10.1093/nar/gkm306>
- 4 UTSUNOMIA, R.; SCACCHETTI, P. C.; PANSONATO-ALVES, J. C.; OLIVEIRA, C.;
5 FORESTI, F. Comparative chromosome mapping of U2 snRNA and 5S rRNA genes in
6 *Gymnotus* Species (Gymnotiformes, Gymnotidae): Evolutionary dynamics and sex
7 chromosome linkage in *G. pantanal*. **Cytogenet Genome Res**, v. 142, p. 286–292, 2014.
8 <https://doi.org/10.1159/000362258>
- 9 VICARI, M. R.; ARTONI, R. F.; MOREIRA-FILHO, O.; BERTOLLO, L. A. C. Basic and
10 molecular cytogenetics in freshwater Cichlidae (Osteichthyes, Perciformes). Karyotypic
11 conservationism and divergence. **Caryologia**, v. 59, p. 260–266, 2006a.
12 <https://doi.org/10.1080/00087114.2006.10797924>
- 13 VICARI, M. R.; MOREIRA-FILHO, O.; ARTONI, R. F.; BERTOLLO, L. A. C. ZZ/ZW sex
14 chromosomes system in an undescribed species of the genus *Apareiodon*
15 (Characiformes, Parodontidae). **Cytogenet Genome Res**, v. 114, p. 163–168, 2006b.
16 <https://doi.org/10.1159/000093333>
- 17 VICARI, M. R.; NOGAROTO, V.; NOLETO, R. B.; CESTARI, M. M.; CIOFFI, M. B.;
18 ALMEIDA, M. C.; MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; ARTONI, R. F. Satellite
19 DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. **J**
20 **Fish Biol**, v. 76, p. 1094–1116, 2010. <https://doi.org/10.1111/j.1095-8649.2010.02564.x>
- 21 VICENTE, V. E.; JESUS, C. M.; MOREIRA-FILHO, O. Chromosomal localization of 5S
22 and 18S rDNA genes in three *Parodon* species (Pisces, Parodontidae). **Caryologia**, v.
23 54, p. 365–369, 2001. <https://doi.org/10.1080/00087114.2001.10589247>
- 24 VICENTE, V. E.; BERTOLLO, L. A. C.; VALENTINI, S. R.; MOREIRA-FILHO, O. Origin
25 and differentiation of sex chromosome system in *Parodon hilarii* (Pisces, Parodontidae).
26 Satellite DNA, G and C-banding. **Genetica**, v. 119, p. 115–120, 2003.
27 <https://doi.org/10.1023/A:1026082904672>
- 28 VISHNOI, A.; RANI, S. MiRNA Biogenesis and Regulation of Diseases: An Overview. *In*:
29 RANI, S. (eds) MicroRNA Profiling. Methods in Molecular Biology, vol 1509. **Humana**
30 **Press**, New York, p. 1–10, 2017. https://doi.org/10.1007/978-1-4939-6524-3_1
- 31 VITALES, D.; D'AMBROSIO, U.; GÁLVEZ, F.; KOVAŘÍK, A.; GARCIA, S. Third release
32 of the plant rDNA database with updated content and information on telomere composition
33 and sequenced plant genomes. **Plant Syst Evol**, v. 303, p. 1115–1121, 2017.
34 <https://doi.org/10.1007/s00606-017-1440-9>
- 35 WALKER, B. J.; ABEEL, T.; SHEA, T.; PRIEST, M.; ABOUELLIEL, A.; SAKTHIKUMAR,
36 S.; CUOMO, C. A.; ZENG, Q.; WORTMAN, J.; YOUNG, S. K.; EARL, A. M. Pilon: An
37 Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly

- 1 Improvement. **PLoS ONE**, v. 9, n. e112963, 2014.
2 <https://doi.org/10.1371/journal.pone.0112963>
- 3 WELLS, J. N.; FESCHOTTE, C. A field guide to eukaryotic transposable elements. **Annu**
4 **Rev Genet**, v. 54, p. 539–561, 2020. [https://doi.org/10.1146/annurev-genet-040620-](https://doi.org/10.1146/annurev-genet-040620-022145)
5 022145
- 6 WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; CAPY, P.; CHALHOUB, B.;
7 FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.;
8 SCHULMAN, H. A unified classification system for eukaryotic transposable elements. **Nat**
9 **Rev Genet**, v. 8, p. 973–982, 2007. <https://doi.org/10.1038/nrg2165>
- 10 WILKINSON, M. E.; CHARENTON, C.; NAGAI, K. RNA splicing by the spliceosome.
11 **Annu Rev Biochem**, v. 89, p. 359–388, 2020. [https://doi.org/10.1146/annurev-biochem-](https://doi.org/10.1146/annurev-biochem-091719-064225)
12 091719-064225
- 13 WILL, C. L.; LÜHRMANN, R. Spliceosome structure and function. **Cold Spring Harb**
14 **Perspect Biol**, v. 3, n. a003707, 2011. <https://doi.org/10.1101/cshperspect.a003707>
- 15 YANO, C. F.; BERTOLLO, L. A. C.; REBORDINOS, L.; MERLO, M. A.; LIEHR, T.;
16 PORTELA-BENS, S.; CIOFFI, M. B. Evolutionary Dynamics of rDNAs and U2 Small
17 Nuclear DNAs in *Triportheus* (Characiformes, Triportheidae): High Variability and
18 Particular Syntenic Organization. **Zebrafish**, v. 14, p. 146–154, 2017.
19 <https://doi.org/10.1089/zeb.2016.1351>
- 20 YANO, C. F.; MERLO, M. A.; PORTELA-BENS, S.; CIOFFI, M. B.; BERTOLLO, L. A. C.;
21 SANTOS-JÚNIOR, C. D.; REBORDINOS, L. Evolutionary Dynamics of Multigene
22 Families in *Triportheus* (Characiformes, Triportheidae): A Transposon Mediated
23 Mechanism? **Front Mar Sci**, v. 7, n. 6, 2020. <https://doi.org/10.3389/fmars.2020.00006>
- 24 YE, C.; CHEN, L.; LIU, C.; ZHU, Q.; FAN, L. Widespread noncoding circular RNAs in
25 plants. **New Phytol**, v. 208, p. 88–95, 2015. <https://doi.org/10.1111/nph.13585>
- 26 ZHANG, Y.; JIANG, W.; GAO, L. Evolution of microRNA genes in *Oryza sativa* and
27 *Arabidopsis thaliana*: an update of the inverted duplication model. **PLoS ONE**, v. 6, n.
28 e28073, 2011. <https://doi.org/10.1371/journal.pone.0028073>
- 29 ZHANG, Y.; YUN, Z.; GONG, L.; QU, H.; DUAN, X.; JIANG, Y.; ZHU, H. Comparasion of
30 miRNA evolution and function in plants and animals. **Microna**, v. 7, p. 4–10, 2018.
31 <https://doi.org/10.2174/2211536607666180126163031>
- 32 ZHANG, P.; WU, W.; CHEN, Q.; CHEN, M. Non-Coding RNAs and their Integrated
33 Networks. **J Integr Bioinform**, v. 16, n. 20190027, 2019. [https://doi.org/10.1515/jib-2019-](https://doi.org/10.1515/jib-2019-0027)
34 0027

- 1 ZIEMNICZAK, K.; TRALDI, J. B.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI, R. F.;
- 2 MOREIRA-FILHO, O.; VICARI, M. R. In situ localization of (GATA)_n and (TTAGGG)_n
- 3 repeated DNAs and W sex chromosome differentiation in Parodontidae (Actinopterygii:
- 4 Characiformes). **Cytogenet Genome Res**, v. 144, p. 325–332, 2014.
- 5 <https://doi.org/10.1159/000370297>