UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS – RIO CLARO

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(BIOLOGIA CELULAR, MOLECULAR E MICROBIOLOGIA)

DESVENDANDO A ARQUITETURA GENÔMICA DE PEIXES *Megaleporinus*: ANÁLISE DE SEQUÊNCIAS REPETITIVAS E SEU IMPACTO NOS SISTEMAS SEXUAIS DAS ESPÉCIES

CAROLINA CREPALDI

Rio Claro – SP
2024

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS – RIO CLARO

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(BIOLOGIA CELULAR, MOLECULAR E MICROBIOLOGIA)

DESVENDANDO A ARQUITETURA GENÔMICA DE PEIXES *Megaleporinus*: ANÁLISE DE SEQUÊNCIAS REPETITIVAS E SEU IMPACTO NOS SISTEMAS SEXUAIS DAS ESPÉCIES

CAROLINA CREPALDI

Tese apresentada ao Instituto de Biociências do Câmpus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Doutor em Ciências Biológicas (Biologia Celular, Molecular e Microbiologia).
Orientador: Drª Patricia Pasquali Parise-Maltempi

Rio Claro – SP
2024

C917d

Crepaldi, Carolina
Desvendando a arquitetura genômica de peixes Megaleporinus: análise de sequências repetitivas e seu impacto nos sistemas sexuais das espécies / Carolina Crepaldi. -- Rio Claro, 2024
72 p. : il., tabs., fotos

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, Rio Claro
Orientadora: Patricia Pasquali Parise-Maltempi

1. Genômica animal. 2. Cromossomos sexuais. 3. Citogenética animal. 4. DNA repetitivo. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DESVENDANDO A ARQUITETURA GENÔMICA DE PEIXES Megaleporinus: ANÁLISE DE SEQUÊNCIAS REPETITIVAS E SEU IMPACTO NOS SISTEMAS SEXUAIS DAS ESPÉCIES

AUTORA: CAROLINA CREPALDI

ORIENTADORA: PATRICIA PASQUALI PARISE MALTEMPI

Aprovada como parte das exigências para obtenção do Título de Doutora em Ciências Biológicas, área: Estrutura, Função e Produção de Biomoléculas pela Comissão Examinadora:

Documento assinado digitalmente
gov.br PATRICIA PASQUALI PARISE MALTEMPI
Data: 09/04/2024 12:49:33-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. PATRICIA PASQUALI PARISE MALTEMPI (Participaçao Virtual)
Departamento de Biologia Geral e Aplicada / Unesp - IB Rio Claro

Prof. Dr. MATEUS MONDIN (Participaçao Virtual)
Departamento de Genética / USP

Documento assinado digitalmente
gov.br DIOGO CAVALCANTI CABRAL DE MELLO
Data: 14/04/2024 20:29:09-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. DIOGO CAVALCANTI CABRAL DE MELLO (Participaçao Virtual)
Departamento de Biologia Geral e Aplicada / Unesp - IB Rio Claro

Prof. Dr. MARCELO RICARDO VICARI (Participaçao Virtual)
Departamento de Biologia Estrutural, Molecular e Genética / Universidade Estadual de Ponta Grossa

Documento assinado digitalmente
gov.br MARCELO RICARDO VICARI
Data: 09/04/2024 15:50:51-0300
Verifique em https://validar.iti.gov.br

Rio Claro, 09 de abril de 2024

AGRADECIMENTOS

Ficam aqui meus mais sinceros agradecimentos a todos que contribuíram de alguma forma com a realização desta tese, com a minha formação como pesquisadora ou com a minha evolução como pessoa. Apesar dos desafios durante este doutorado, passando por uma pandemia por praticamente metade do período e pela infelicidade do incêndio no antigo Departamento de Biologia e da perda do nosso laboratório, entrego este trabalho sabendo que a ajuda dos que caminharam comigo foi essencial para esta fase.

À toda a minha família que sempre me deu amor incondicional, especialmente meus pais, Regina e Luiz, por todo o apoio, cuidado e ternura. Vocês são maravilhosos em tantos aspectos. Hoje só é possível me chamar de cientista graças a vocês dois. Obrigada por tudo!

Ao meu namorado, companheiro de vida e melhor amigo, Caio Zamuner. Nós caminhamos unidos já faz tanto tempo e cada dia que passa nosso amor só transborda ainda mais. Somos tão fortes juntos e prestes a vencer mais esta etapa juntos também. Te amo!

Aos meus melhores amigos e irmãos de formação do CB 2011, desde os que convivo todos os dias até os que estão um pouco mais longe. Sempre fomos vistos como uma grande família por muitas pessoas e isso é a mais completa verdade. A companhia e amizade de vocês foi e sempre será fundamental, levo todos vocês sempre juntos comigo.

À minha orientadora e exemplo de mulher forte, Prof.ª Dr.ª Patricia Pasquali Parise-Maltempi. Obrigada pelo carinho e dedicação durante todos esses anos trabalhando juntas, e pela confiança depositado em mim em mais uma etapa da minha formação.

Ao pessoal do nosso laboratório que caminha comigo já há muitos anos, em especial Luiza, Marcelo, Murilo, Gabriel, Evélin, vocês são pessoas maravilhosas. Obrigada a vocês e tantos outros amigos que conviveram comigo por anos debaixo do mesmo teto científico.

Aos integrantes do Departamento de Biologia Geral e Aplicada, em especial ao Prof. Diogo Cabral e seu grupo e aos alunos dos laboratórios de microbiologia, pelo convívio diário e pelas amizades tanto novas quanto mais antigas.

À Universidade Estadual Paulista “Júlio de Mesquita Filho”, campus de Rio Claro, e ao Programa de Pós-Graduação em Ciências Biológicas - Biologia Celular e Molecular.

Ao financiamento de processo nº 2017/00195-7, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“Existe uma teoria que diz que, se um dia alguém descobrir exatamente para que serve o Universo e por que ele está aqui, ele desaparecerá instantaneamente e será substituído por algo ainda mais estranho e inexplicável. Existe uma segunda teoria que diz que isso já aconteceu.”

Douglas Adams, O Guia do Mochileiro das Galáxias.

RESUMO

A caracterização da porção repetitiva do material genético pode trazer informações sobre a composição e a estrutura genômica de uma espécie, bem como sobre os mecanismos envolvidos na evolução de seu genoma e no aparecimento de inovações genéticas, como os cromossomos sexuais, por exemplo. Uma característica intrínseca à evolução de cromossomos sexuais é o acúmulo de DNAs repetitivos, como DNAs satélites (satDNAs) e elementos transponíveis (TEs), envolvidos em sua diferenciação e diversificação em vertebrados. Cromossomos sexuais se apresentam como entidades dinâmicas dentro de um genoma, com evolução, morfologia e conteúdo genéticos únicos. Estes elementos específicos evoluíram de maneiras independentes dentro de linhagens de vertebrados, se apresentando nos mais diversos sistemas, especialmente em peixes, anfíbios e répteis. O presente trabalho teve como objetivo buscar descrever o conteúdo repetitivo em espécies de peixes neotropicais relacionados com diferentes sistemas de cromossomos sexuais: *Megaleporinus elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$), *Megaleporinus macrocephalus* (ZZ/ZW) e *Leporinus friderici* (sem cromossomos sexuais heteromórficos). Assim, nosso foco foi investigar as diferenças interespecíficas destas sequências, avaliando suas possíveis contribuições para a evolução genômica e de cromossomos sexuais. Para isto, associamos o uso de plataformas de Sequenciamento de Próxima Geração (NGS) com análises de bioinformática para um acesso mais refinado ao genoma das espécies, permitindo a anotação destes DNAs repetitivos, avaliação de suas dinâmicas e evolução, bem como a distribuição física de satDNAs nos cromossomos. Primeiramente caracterizamos e descrevemos o satelitoma (conjunto de famílias de satDNAs) de *Megaleporinus elongatus* (CAPÍTULO 1), onde comparamos as composições de satDNAs entre os sexos da espécie. Constatamos que apesar de possuir o cromossomo sexual altamente diferenciado e heterocromático, a fêmea não apresenta uma diversidade de famílias tão maior do que o macho – ambos possuem uma alta diversidade de famílias enviesadas para seus genomas em comparação ao outro sexo. Constatamos então que foi o acúmulo e a expansão de apenas algumas famílias mais abundantes na fêmea que levou à alta diferenciação do cromossomo sexual, e não necessariamente a presença de uma grande diversidade de satDNAs diferentes ancorados a esse cromossomo. A história evolutiva das duas sequências mais abundantes também dita como suas cópias se expandiram de maneiras diferentes entre *M. elongatus* e *M. macrocephalus*. Estes dados corroboraram a hipótese de composições diferentes e evolução espécie-específica para satDNAs, mesmo entre espécies próximas, e, neste caso, com cromossomos sexuais heteromórficos. Em seguida, avaliamos a composição de elementos transponíveis (TEs)

(CAPÍTULO 2) para a descrição e a comparação dos mobilomas de fêmeas de *M. elongatus* e *M. macrocephalus* juntamente de uma terceira espécie, *Leporinus friderici*, completando três sistemas sexuais diferentes entre si. Concluímos que apesar da diversidade de famílias de TEs ser relativamente semelhante entre os três genomas, as dinâmicas de invasão e amplificação destas sequências também são espécie-específicas. Eventos de expansão foram particulares de cada genoma estudado, e a expansão rápida e massiva de TEs ocorreu em um período evolutivo curto, especialmente após a separação das espécies e o surgimento dos cromossomos sexuais em *Megaleporinus*. Os dados obtidos nesta tese fornecem mais informações sobre as características e a organização de satDNAs e transposons e suas respectivas dinâmicas nas espécies estudadas, bem como sua distribuição, incidência e influência ao longo da evolução dos cromossomos sexuais do grupo.

Palavras-chave: Citogenômica; elementos repetitivos; elementos transponíveis; satDNAs; cromossomos sexuais heteromórficos.

ABSTRACT

The characterization of the repetitive portion of the generic material can provide information about the composition and genomic structure of a species, as well as the mechanisms involved in the evolution of its genome and the emergence of genetic innovations, such as sexual chromosomes. An intrinsic feature in the evolution of sexual chromosomes is the accumulation of these repetitive DNAs, such as satellite DNAs (satDNAs) and transposable elements (TEs), which are involved in their differentiation and diversification in vertebrates. Sex chromosomes present themselves as dynamic entities within a genome, with unique evolution, morphology, and genetic content. These specific elements have evolved independently within vertebrate lineages, appearing in various systems, especially in fish, amphibians, and reptiles. The present work aimed to describe the repetitive content in related neotropical fish species with different sexual chromosome systems: *Megaleporinus elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$), *Megaleporinus macrocephalus* (ZZ/ZW), and *Leporinus friderici* (without heteromorphic sexual chromosomes). Thus, our focus was to investigate interspecific differences in these sequences, evaluating their possible contributions to genomic and sexual chromosome evolution. To achieve this, we employed Next-Generation Sequencing (NGS) platforms along with bioinformatics analyses for a more refined access to the species' genomes, allowing the annotation of these repetitive DNAs, assessment of their dynamics and evolution, as well as the physical distribution of satDNAs on chromosomes. Firstly, we characterized and described the satellitome (the whole set of satDNA families) of *Megaleporinus elongatus* (CHAPTER 1), comparing the satDNA composition between the sexes. Despite having a highly differentiated and heterochromatic sex chromosome, the female does not exhibit a significantly greater diversity of families than the male – both have equally high diversity of sex-biased satDNAs families. We then observed that the accumulation and expansion of only a few more abundant families in the female led to the extreme differentiation of the sex chromosome, not necessarily the presence of a larger diversity of different satDNAs anchored to this chromosome. The evolutionary history of the two most abundant sequences also dictated how their copies expanded differently between *M. elongatus* and *M. macrocephalus*. These data supported the hypothesis of different compositions and species-specific evolution for satDNAs, even among closely related species, and, in this case, with heteromorphic sex chromosomes. Next, we evaluated the composition of transposable elements (TEs) (CHAPTER 2) for the description and comparison of the mobilomes of *M. elongatus* and *M. macrocephalus* females, along with a third species, *Leporinus friderici*, representing three distinct sex chromosome systems. We

concluded that despite the relatively similar diversity of TE families among the three genomes, the dynamics of invasion and amplification of these sequences are also species-specific. Expansion events are unique to each studied genome, and the rapid and massive expansion of TEs occurred in a short evolutionary period, especially after the separation of species and the emergence of sexual chromosomes in *Megaleporinus*. The data obtained in this thesis provide more information about the characteristics and organization of satDNAs and transposons and their respective dynamics in the studied species, as well as their distribution, incidence, and influence throughout the evolution of the sex chromosomes in the group.

Keywords: Cytogenomics; repetitive elements; transposable elements; satDNAs; genome evolution; heteromorphic sex chromosomes.

SUMÁRIO

1 INTRODUÇÃO GERAL

1.1 Considerações sobre cromossomos sexuais e sua ocorrência em peixes

A determinação sexual é um processo vital para vertebrados, uma vez que dita quais indivíduos irão se desenvolver em machos e quais serão fêmeas. Muitas espécies apresentam determinação genética do sexo, no qual genes determinantes de características sexuais ou supressoras de características do sexo oposto irão estabelecer qual será o sexo do embrião em desenvolvimento. Os cromossomos que adquirem estes genes determinantes são denominados cromossomos sexuais, tornando-se as entidades genômicas responsáveis pela determinação do sexo e se diferenciando de seu par e do restante do conjunto cromossômico do indivíduo (Charlesworth, 1991; Bachtrog, 2006; Graves, 2008; Bergero and Charlesworth, 2009; Wright *et al.*, 2016; Charlesworth, 2021b). As origens e parâmetros de evolução de cromossomos sexuais ainda são alvos de pesquisas atuais uma vez que se mostram como elementos altamente dinâmicos, constituindo bons modelos para estudos evolutivos, genômicos e comparativos.

Os sistemas de cromossomos sexuais geralmente são categorizados em heterogametias feminina e masculina, caracterizados pela presença de cromossomos sexuais W nas fêmeas (sistema ZW) e de cromossomos sexuais Y (sistema XY) nos machos, respectivamente. Estes cromossomos, no entanto, podem assumir as mais diversas configurações, com sua identidade se mantendo ou se alterando através de diversos mecanismos evolutivos (Figura 1) (Wright *et al.*, 2016; Palmer *et al.*, 2019; Furman *et al.*, 2020).

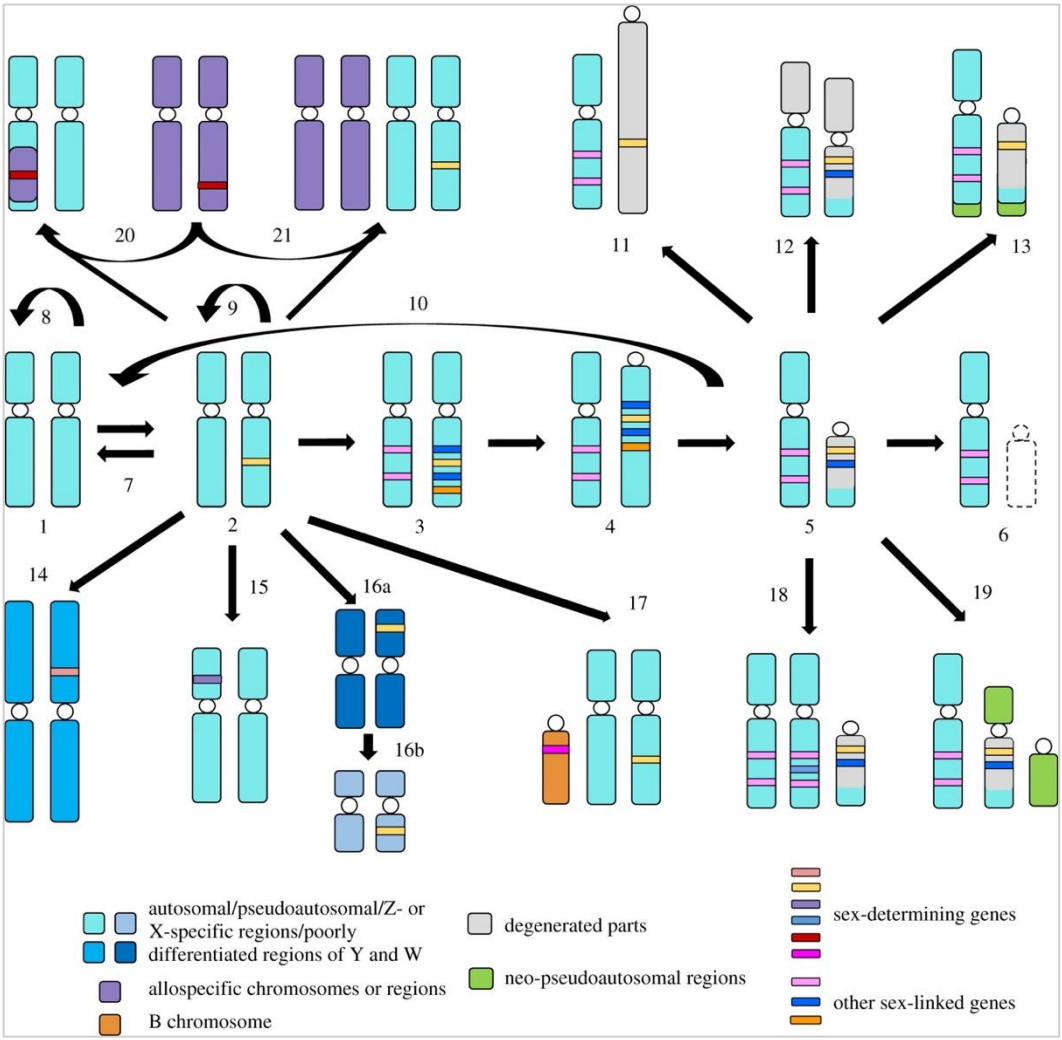


Figura 1. Visão geral dos passos na evolução dos cromossomos sexuais em vertebrados. Esta rede hipotética de trajetórias evolutivas pode ramificar-se em resultados potencialmente infinitos, com a possibilidade de estagnar por um longo período e até mesmo reverter para certos estados. Os passos (1-6) são os eventos considerados canônicos, mas genes sexualmente antagonistas não estão necessariamente envolvidos e podem ser substituídos aqui por genes ligados ao sexo em geral; (7) - mudança para hermafroditismo ou ESD, onde não há cromossomos sexuais presentes; (8) - evolução de longo prazo sem emergência de cromossomos sexuais; (9) - persistência de longo prazo de cromossomos sexuais pouco diferenciados; (10) - reversão para estágios com cromossomos sexuais menos diferenciados; (11) - expansão de repetições no cromossomo sexual específico do sexo causando sua mudança de tamanho; (12) - acumulação de repetições em ambos os cromossomos sexuais; (13) - fusão dos cromossomos sexuais com um autossomo levando à expansão da região pseudoautossomal; (14) - emergência de um novo *locus* de determinação sexual em outro cromossomo; (15) - emergência de um novo *locus* de determinação sexual dentro dos cromossomos sexuais existentes; (16a, b) - duas translocações do mesmo *locus* de determinação sexual para outros cromossomos; (17) - origem de um novo sistema de determinação sexual com envolvimento de cromossomos B; (18) - emergência de sistemas de determinação sexual com três cromossomos sexuais homólogos; (19) - fusão de cromossomos sexuais com um autossomo levando a

múltiplos neo-cromossomos sexuais; (20) - introgressão de um gene determinante do sexo de uma população ou espécie diferente; (21) - alopoliploidização conectada com a emergência de um novo sistema de determinação sexual em um genoma de origem híbrida. Para simplificação, apenas o sexo heterogamético é representado. Fonte: Kratochvíl *et al.* (2021).

Assim, a evolução de um cromossomo sexual não segue necessariamente um processo unilateral de diferenciação. Este cromossomo especializado pode se manter idêntico ou se alterar completamente de seu par; pode sofrer degeneração completa; ou até mesmo ser trocado por um novo par cromossômico que se torna o novo par sexual (evento denominado de *turnover*) (Bachtrog, 2006; Graves, 2008; Wright *et al.*, 2016; Furman *et al.*, 2020) (Figura 1, passo 14).

Esses diversos eventos evolutivos e configurações de cromossomos sexuais se refletem igualmente na diversidade de ocorrência deles ao longo dos grupos de eucariotos (Figura 2). Embora tenham funcionalidade universal no caso, a determinação do sexo cromossomos sexuais são bons exemplos de evolução convergente à nível genômico, no qual embora apresentem funcionalidade semelhante, estes evoluíram repetidas vezes, de maneira independente e nas mais diversas configurações ao longo dos organismos.

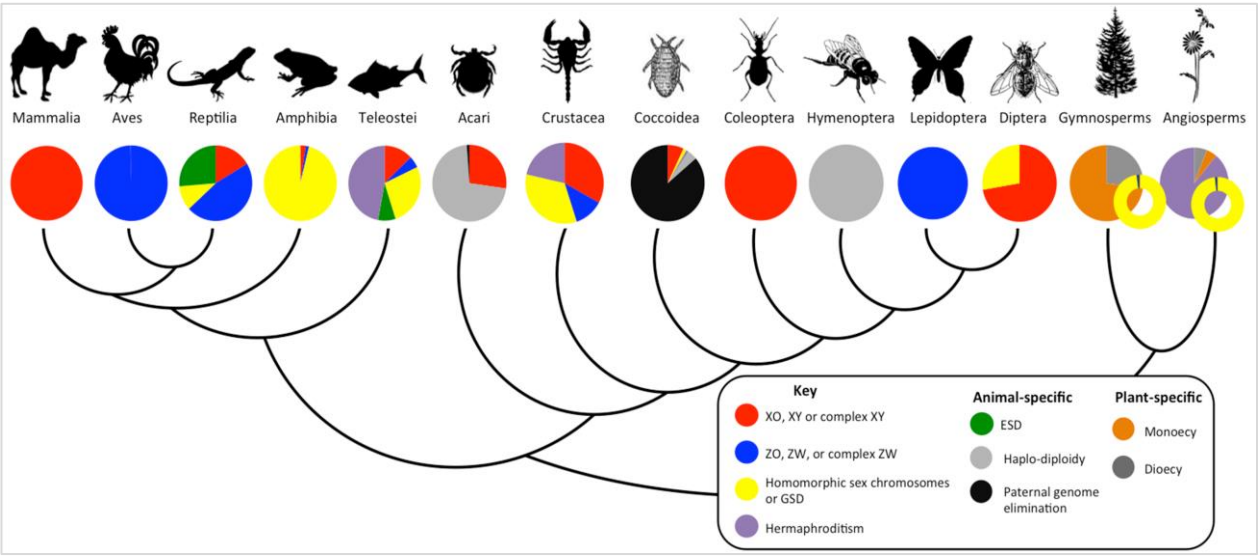


Figura 2. Modelo esquemático para os tipos de sistemas de cromossomos sexuais encontrados em diversos grupos eucarióticos. Na legenda estão descriminados os sistemas de cromossomos sexuais mais comuns, incluindo sistemas de heterogametia masculina (machos com o cromossomo sexual, sistemas XO, XY), heterogametia feminina (fêmeas com cromossomos sexuais, sistemas ZO, ZW), determinação genética do sexo sem cromossomos sexuais diferenciados (GSD – *genetic sex determination*), determinação ambiental do sexo (ESD – *environmental sex determination*), entre outros. Atente para a diversidade de sistemas

encontrada dentro do grupo dos vertebrados à esquerda da ilustração. Fonte: Bachtrog *et al.* (2014).

Dentro do grupo dos vertebrados já é possível ver essa diversidade significativa entre os principais grupos, visto que evoluíram repetidas vezes e nas mais diversas configurações (Bachtrog, 2006; Bachtrog *et al.*, 2011; Furman *et al.*, 2020; Kratochvíl *et al.*, 2021) (Figura 2). Mamíferos e aves, por exemplo, apresentam sistemas antigos, conservados e bem definidos, com heterogametias masculina XY e feminina ZW, respectivamente; já peixes, répteis e anfíbios exibem uma miríade de sistemas, com surgimentos independentes, mais recentes e muito rápidos.

Em se tratando de peixes, especificamente, os cromossomos sexuais podem estar configurados em sistemas simples ou múltiplos, com cromossomos sexuais sem diferenciação morfológica de seu par (homomórficos) ou altamente diferenciados (Figura 2). A ocorrência de mais de um cromossomo sexual, ditando um sistema sexual múltiplo, por exemplo, segue inclusive outros padrões evolutivos do convencional para a diferenciação de um cromossomo heteromórfico. O surgimento constante e individual de sistemas múltiplos e simples em peixes é muito comum, especialmente considerando os mecanismos variáveis e transitórios de determinação sexual genética no grupo (Parise-Maltempi *et al.*, 2007; Kitano and Peichel, 2012; Blanco *et al.*, 2014; Bueno *et al.*, 2015; De Oliveira *et al.*, 2018).

Há também exemplos de transições entre as heterogametias de machos e fêmeas e frequentes eventos de *turnover* (o recrutamento de um novo par para ser o determinante do sexo) (Graves, 2008; Bachtrog *et al.*, 2011; Charlesworth, 2019; Furman *et al.*, 2020; Kratochvíl *et al.*, 2021) (Figura 1, passo 14). Este *turnover* inclusive é um dos aspectos que dita essa variedade de sistemas sexuais em peixes, além dessa ocorrência de sistemas de cromossomos sexuais mais jovens e recentes em termos evolutivos em relação aos sistemas mais antigos e conservados de mamíferos e aves (Kitano and Peichel, 2012; Myosho *et al.*, 2015; Schartl *et al.*, 2016; Charlesworth, 2019). Com isso, a aparição de novos cromossomos sexuais em teleósteos pode ocorrer repetidas vezes, acarretando em sistemas distintos entre espécies próximas e até populações distintas (Mank *et al.*, 2006; Mank and Avise, 2009; Chalopin *et al.*, 2015b; Schartl *et al.*, 2016) (Figura 3), como é o caso do gênero *Triportheus*, onde cada espécie apresenta tamanhos distintos de cromossomos W entre si (Yano *et al.*, 2016a; Cioffi *et al.*, 2017) (Figura 3a), e de diversos indivíduos do gênero *Oryzias*, com diferentes

sistemas de cromossomos sexuais e inclusive diferentes genes determinantes do sexo entre espécies próximas (Gammerdinger and Kocher, 2018) (Figura 3b).

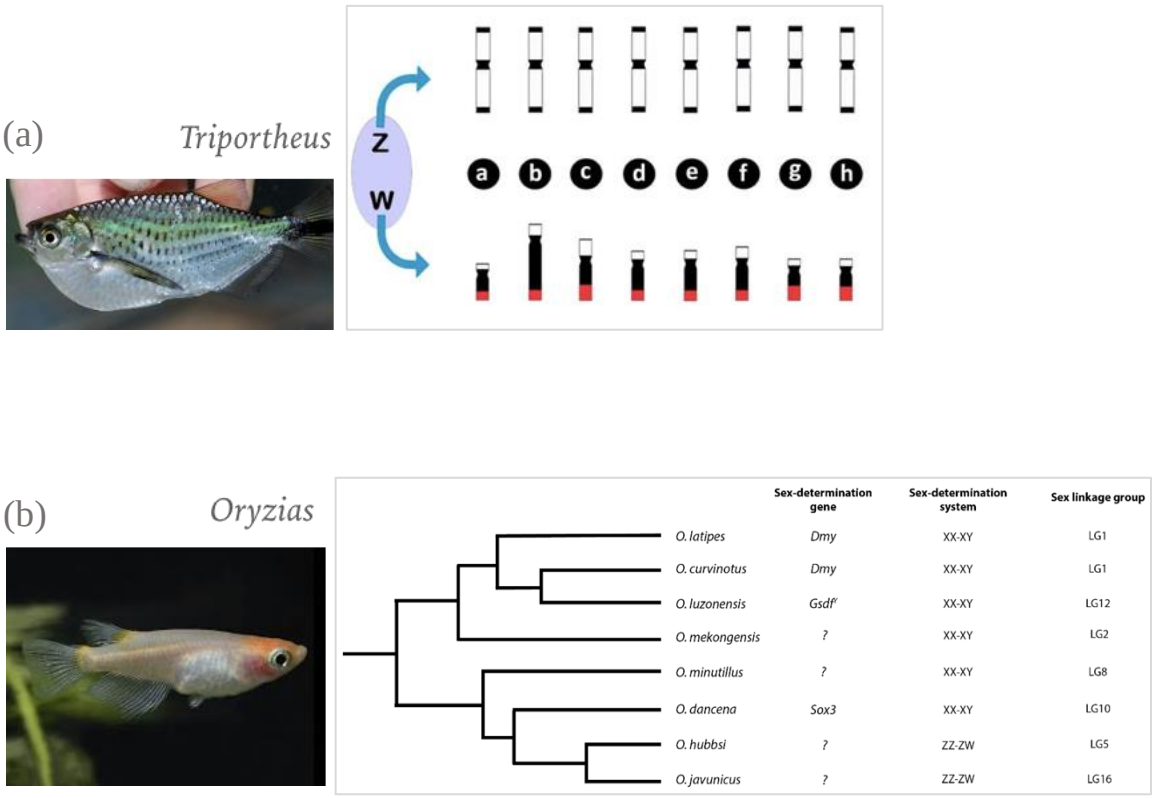


Figura 3. Exemplos esquemáticos de gêneros distintos de peixes teleósteos e suas respectivas variedades de sistemas de cromossomos sexuais, tanto em diferenças de morfologia do cromossomo sexual quanto em diferentes sistemas ocorrendo entre espécies próximas. Gênero *Triportheus* (a) com modelo esquemático de cromossomos sexuais W de diferentes tamanhos para fêmeas de diferentes espécies (a-h). Fonte: foto de J. Grad (PyBio Paraguay Biodiversidad); esquematização dos cromossomos sexuais adaptada de Cioffi *et al.* (2017). Gênero *Oryzias* (b) com ocorrência de diferentes genes de determinação sexual (*sex-determination gene*), diferentes sistemas de heterogametia (*sex-determination system*) para diferentes espécies. Fonte: foto de diapteron.co.uk; filogenia e dados de determinação sexual de Gammerdinger and Kocher (2018).

1.2 Cromossomos sexuais heteromórficos e seu conteúdo repetitivo

O entendimento dos processos evolutivos que regem a formação dos sistemas sexuais em peixes ainda permanece limitado, especialmente devido à diversidade de espécies e à alta labilidade dos cromossomos no grupo. A variação de sistemas sexuais é geralmente atribuída à alta capacidade de *turnover* destes cromossomos no grupo (Kitano and Peichel, 2012; Yoshida

et al., 2014; Charlesworth, 2019; Palmer *et al.*, 2019) e também ao dinamismo de DNAs repetitivos em seus genomas, muitas vezes acumulados nos cromossomos sexuais (Chalopin *et al.*, 2015b; Charlesworth, 2019). O acúmulo de DNAs repetitivos inclusive pode proporcionar diferenças na estrutura e na composição destes cromossomos, o que também ocasiona a diferenciação entre o sexual com seu par, gerando os chamados cromossomos sexuais heteromórficos (como exemplo demonstrado acima, os cromossomos W com morfologias discrepantes tanto entre si quanto com seus pares Z em espécies de *Triportheus* (Figura 3a)). Esta diferenciação de forma geralmente ocorre por supressão de recombinação entre o novo cromossomo sexual e seu par, o que gera um consequente acúmulo de DNA não-codificante (heterocromatina) em especial em sistemas de cromossomos sexuais simples. Por consequência, cromossomos sexuais podem então se tornar heteromórficos (diferentes de seus pares) tanto em conteúdo genético quanto em estrutura com o acúmulo massivo destes DNAs repetitivos (Bergero and Charlesworth, 2009; Wright *et al.*, 2016; Furman *et al.*, 2020; Charlesworth, 2021a).

DNAs repetitivos são sequências de DNA que, diferentes dos genes e das sequências determinantes do sexo por exemplo, são não-codificantes e podem se encontrar repetidas vezes dentro do genoma eucariótico. Assim, um cromossomo sexual nascente pode adquirir, além da nova característica sexual, esse acúmulo de sequências repetidas não-codificantes o que pode vir mudar a sua forma, tornando-o discrepante do seu par. Esse DNA repetitivo acumulado pode estar distribuído ao longo de toda a estrutura do cromossomo sexual, até estarem completamente tomados, ou presentes em apenas partes deles, denotando mais um grau de variedade para cromossomos sexuais no grupo.

DNAs repetitivos são intrínsecos de genomas de eucariotos e podem estar presentes nas mais diversas abundâncias e configurações, inclusive nos cromossomos sexuais heteromórficos. Essencialmente, genomas eucarióticos apresentam sequências nucleotídicas divididas essencialmente em dois grandes grupos: regiões gênicas, com cópias únicas codificantes de genes funcionais; e estes DNAs repetitivos, regiões do material genético com sequências presentes em diversas cópias, repetidas inúmeras vezes ao longo do DNA, com tamanhos variáveis e correspondendo a grandes frações genômicas (Charlesworth *et al.*, 1994; Plohl *et al.*, 2012a; Garrido-Ramos, 2017).

Apesar de majoritariamente não-codificantes, DNAs repetitivos são responsáveis por estrutura, tamanho e variação genômica entre espécies (Meštrović *et al.*, 2015; Garrido-Ramos, 2017; Jagannathan *et al.*, 2018). Estas sequências podem também se inserir em locais de genes

ativos, participar da formação e manutenção de regiões centroméricas, levar à diferenciação de cromossomos sexuais e até contribuir com a especiação (Makałowski, 2000; Plohl *et al.*, 2008; Pezer *et al.*, 2012; Plohl *et al.*, 2014; Schartl *et al.*, 2016; Jagannathan *et al.*, 2018; Serrato-Capuchina and Matute, 2018; Camacho *et al.*, 2021b).

São sequências que estão frequentemente submetidas à diversos processos de amplificação, deleção e mutações, sendo altamente variáveis e podendo evoluir também de maneira independente entre espécies (Ugarković and Plohl, 2002; Lorite *et al.*, 2004; Feschotte and Pritham, 2007; Lorite *et al.*, 2017; Utsunomia *et al.*, 2017; Shao *et al.*, 2019). Assim, DNAs repetitivos possuem uma notória variabilidade no que concerne sua ocorrência, abundância, acúmulo e composição molecular (Kejnovsky *et al.*, 2009; Garrido-Ramos, 2017; Šatović-Vukšić and Plohl, 2023).

DNAs repetitivos geralmente são encontrados em sequências altamente específicas e repetidas *in tandem*, como DNAs satélites (satDNAs), ou em sequências dispersas como as diversas famílias de elementos transponíveis (TEs), sejam de transposons de DNA ou retrotransposons (Figura 4).

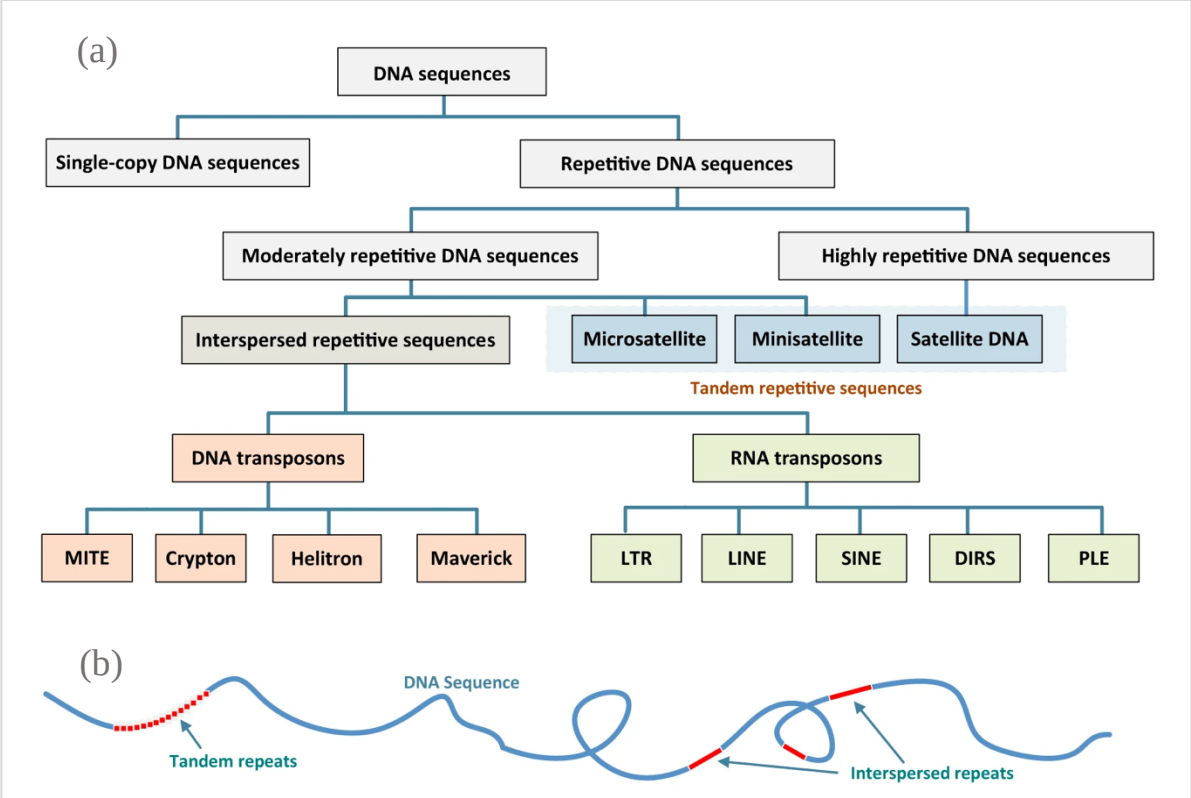


Figura 4. Classificação geral de sequências de DNAs repetitivos no material genético (a), com a organização usual dos dois grandes tipos de repetição de cópias (elementos repetidos *in tandem* e elementos dispersos) em um genoma (b). Fonte: adaptado de Liao *et al.* (2023).

O grupo de DNAs repetitivos classificado como DNAs satélite (satDNAs), constitui uma fração significativa do genoma de animais, peixes teleósteos especialmente, com unidades de repetição, ou monômeros, de tamanhos variáveis e organizados repetidas vezes *in tandem* (Vicari *et al.*, 2010; Schemberger *et al.*, 2011; Garrido-Ramos, 2017; Camacho *et al.*, 2021a) (Figura 4b). São sequências que podem formar longos arranjos de dezenas a milhares de nucleotídeos, sendo tipicamente encontradas em regiões heterocromáticas como centrômeros, telômeros, regiões pericentroméricas (Jagannathan *et al.*, 2018; Rodrigues *et al.*, 2019; Suntronpong *et al.*, 2020; Thakur *et al.*, 2021; Ugarković, 2021) ou outras ocorrências de acúmulo de heterocromatina, como os cromossomos sexuais diferenciados e cromossomos B (Yano *et al.*, 2016b; Crepaldi and Parise-Maltempi, 2020; Stornioli *et al.*, 2021).

O caminho evolutivo de uma família de satDNA se dá a partir do nascimento e duplicação de uma sequência preexistente no genoma. Estas muitas vezes são provenientes de um elemento transponível e podem formam arranjos (*arrays*) de sequências repetidas, que se amplificam de maneira massiva ao serem disseminadas pelo genoma (Meštrović *et al.*, 2015;

Ruiz-Ruano *et al.*, 2016; Ruiz-Ruano *et al.*, 2017; Paço *et al.*, 2019). As dinâmicas evolutivas de sequências de satDNAs costumam envolver a chamada hipótese da biblioteca (Fry and Salser, 1977; Meštrović *et al.*, 2006), que sugere que espécies próximas compartilham satDNAs provenientes de um ancestral comum; estes, no entanto, podem sofrer pressões seletivas diferentes ao longo da evolução das espécies e resultar em graus de amplificação, fixação e localização distintos (Dover, 1986; Dover *et al.*, 1993; Meštrović *et al.*, 2006; Plohl *et al.*, 2012b; Ruiz-Ruano *et al.*, 2016).

O compartilhamento de sequências de satDNA é uma ocorrência que já foi documentada em peixes (Schemberger *et al.*, 2011; Silva *et al.*, 2017; Utsunomia *et al.*, 2017; Crepaldi *et al.*, 2021; Dos Santos *et al.*, 2021; Goes *et al.*, 2023), mostrando que os dois aspectos não excludentes de sequências de DNAs satélites: um alto grau de diversidade de nucleotídeos em suas sequências que estão sob frequente pressão seletiva e podem mudar rapidamente, mas também a ocorrência compartilhada de sequências iguais entre diferentes espécies. Assim, diferentes composições repetitivas podem ditar perfis espécie-específicos, onde diferem no que concerne a sua dinâmica de expansão e abundância dentro de um genoma, apesar da origem compartilhada (Ugarković and Plohl, 2002; Utsunomia *et al.*, 2017; Sember *et al.*, 2018).

Elementos transponíveis (TEs) ou transposons, por sua vez, são sequências de DNAs repetitivos que podem se mover e se inserir em locais diferentes dentro de um genoma. Estão presentes todos os genomas conhecidos até o momento e costumam compreender grandes porções do material genético com sequências grandes, podendo chegar até milhares de pares de base dispersas em múltiplas cópias (Feschotte and Pritham, 2007). São componentes intrínsecos de genomas de teleósteos, e podem contribuir significantemente para a estrutura e tamanho genômicos de peixes (Biémont and Vieira, 2006; Wicker *et al.*, 2007; Carducci *et al.*, 2020), influenciando a dinâmica evolutiva de um genoma tanto a curto quanto a longo prazo (Ferreira *et al.*, 2011; Belyayev, 2014; Carducci *et al.*, 2018; Hayward and Gilbert, 2022).

TEs possuem uma grande diversidade de sequências e elementos que os compõem, sendo tipicamente classificados com base em seus mecanismos de replicação. Eles podem disseminar suas cópias utilizando transcriptases reversas (RT) e intermediários de RNA para se copiarem em novas localizações, caracterizando TEs Classe I, ou retrotransposons; ou utilizando transposases (TPase) e sem nenhum RNA intermediário, operando diretamente ao nível de DNA e sendo retirado e recolocado no genoma, compreendendo TEs Classe II ou transposons de DNA (Wicker *et al.*, 2007; Bourque *et al.*, 2018) (Figura 4a). Essas duas classes amplas ainda podem ser desmembradas em subclasses, superfamílias e famílias que diferem em

estrutura, arranjo e intermediários produzidos, como as subblasses de retrostransposons LTRs (*long terminal repeats*) e non-LTRs, cada uma com diferentes representantes específicos, e as subclasses polintrons, helitrons, cryptons e transposons que utilizando transposase (hats e marienrs) para transposons de DNA (Feschotte and Pritham, 2007; Shao *et al.*, 2019).

Devido à sua diversidade e natureza dinâmica, transposons são grandes candidatos a análises genômicas, especialmente em estudos de espécies não-modelo (Makałowski *et al.*, 2019; Etchegaray *et al.*, 2022). A anotação de bibliotecas espécie-específicas de TEs e de suas características é uma ferramenta útil para o entendimento da evolução destas sequências e seu consequente impacto em genomas de peixes, já tendo sido atrelados à evolução de cromossomos sexuais (Chen *et al.*, 2014; Chalopin *et al.*, 2015b; Schemberger *et al.*, 2019), cromossomos B (Coan and Martins, 2018), rearranjos cromossômicos (Rodríguez *et al.*, 2021) e até comportamentos migratórios (Carotti *et al.*, 2021) dentro do grupo. A análise de diferenças em variedade, abundância e histórico de proliferação entre genomas, por exemplo, podem trazer respostas aos fatores que regem a dinâmica evolutiva destas sequências e de seus hospedeiros (Feschotte and Pritham, 2007; Shao *et al.*, 2019; O'neill *et al.*, 2020; Carotti *et al.*, 2021; Colonna Romano and Fanti, 2022).

DNAs repetitivos são altamente difundidos em vertebrados, com sua dinâmica de frequentes amplificações e deleções contribuindo significantemente para a variação de tamanhos e diversidade genômica (Plohl *et al.*, 2012a; Almojil *et al.*, 2021; Colonna Romano and Fanti, 2022).

No caso dos peixes, informações sobre a organização e a evolução genômica como um todo ainda são limitadas a poucas espécies em comparação à diversidade do grupo. É sabido, no entanto, que apresentam a maior variedade de elementos repetitivos, especialmente transponíveis, em comparação a outros vertebrados (Chalopin *et al.*, 2015a). A proporção genômica que compreende TEs em peixes, por exemplo, é bastante variável , correspondendo a menos de 10% do genomas de tetraodon e fugu e mais de 50% do genoma de zebrafish (Chalopin *et al.*, 2015a) (Figura 5), e costumam estar presentes em diferentes famílias com tamanhos, localização, abundância e constituições diferentes, (Garrido-Ramos, 2017; Thakur *et al.*, 2021).

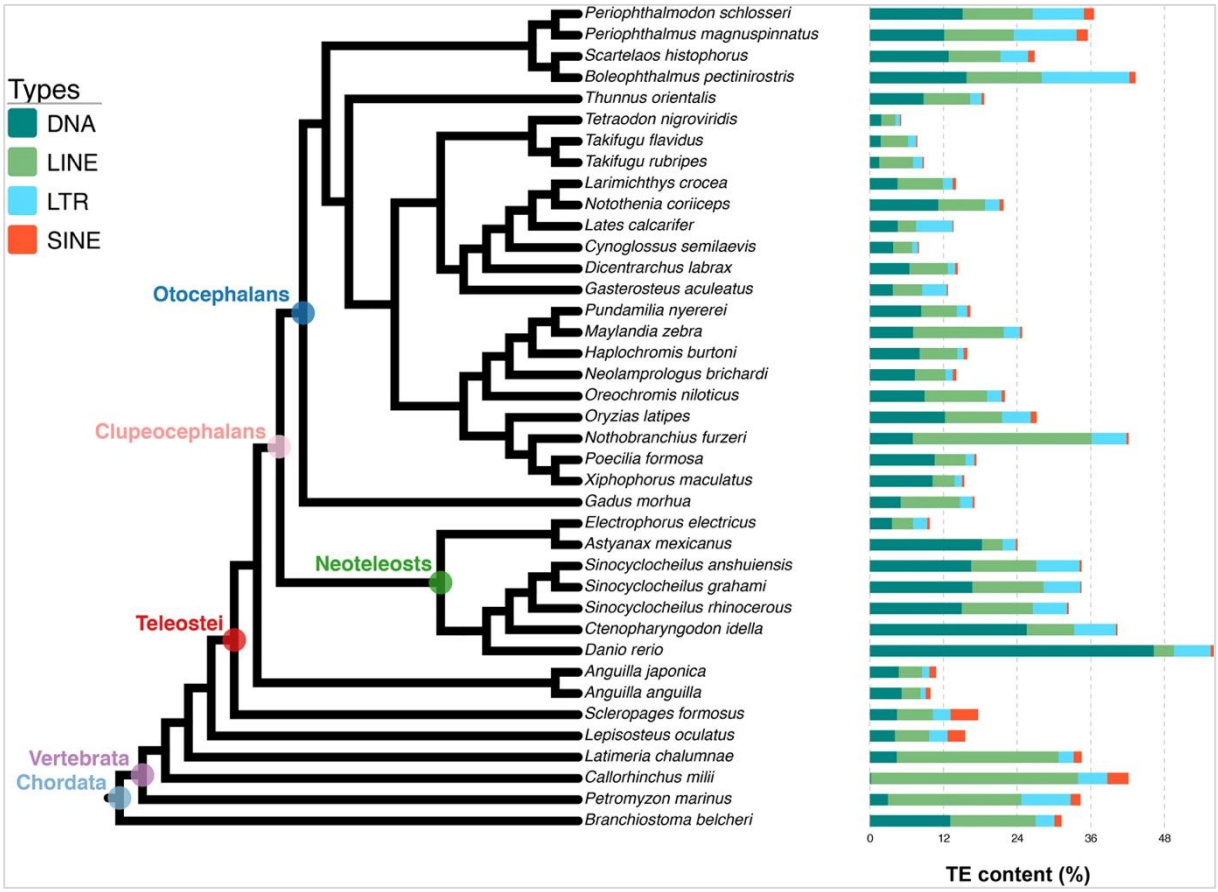


Figura 5. Diversidade de tipos e proporção genômica (%) de elementos transponíveis (transposons de DNA e retrotransposons) em genomas de diferentes espécies de peixes (anfioxo inclusive). Fonte: Shao *et al.* (2019).

Assim, tem-se crescido o interesse na busca, descrição e análise de DNAs repetitivos justamente por sua natureza intrínseca à evolução genômica do grupo, sendo considerados úteis em estudos de evolução cromossômica, evolução de cromossomos sexuais, comparação e caracterização de espécies. O avanço em análises de dados de sequenciamento, atrelado às ferramentas bioinformáticas cada vez mais acessíveis e refinadas, garantiram maior eficiência na detecção destas sequências, incluindo seus padrões de origem, dinâmica e evolução em genomas de espécies não-modelo.

1.3 Os sistemas de cromossomos sexuais e seus conteúdos repetitivos no grupo de peixes *Megaleporinus*

Considerando os dois pontos principais levantados acima para o grupo dos peixes, como a alta variedades de composições genômicas, especialmente de elementos repetitivos, e uma alta diversidade de sistemas de cromossomos sexuais, o foco da presente tese foram os peixes dos gêneros *Megaleporinus* e *Leporinus*, conhecidos popularmente como piaus ou piaparas (Birindelli and Britski, 2013; Eschmeyer and Fong, 2015) (Figura 6). Um dos aspectos mais relevantes desses dois gêneros é a estrutura conservada de 2n = 54 cromossomos entre eles, com morfologias muito similares dos seus pares cromossômicos, com exceção da presença de um par de cromossomos sexuais heteromórficos altamente diferenciados nas fêmeas de *Megaleporinus* e que estão ausentes em *Leporinus* (Parise-Maltempi *et al.*, 2007; Da Silva *et al.*, 2012; Marreta *et al.*, 2012; Ramirez *et al.*, 2017b) (Figura 6a). Esta heterogametia feminina em *Megaleporinus* apresenta cromossomos W como os maiores do cariótipo, muito diferenciado e com uma alta taxa de heterocromatina e DNA repetitivo especialmente nos braços longos. A ocorrência destes cromossomos sexuais foi inclusive uma das características que levou ao desmembramento das espécies nos dois gêneros, uma vez que todas eram previamente classificadas como *Leporinus* (Galetti Jr and Foresti, 1986; Ramirez *et al.*, 2017b) (Figura 6b).

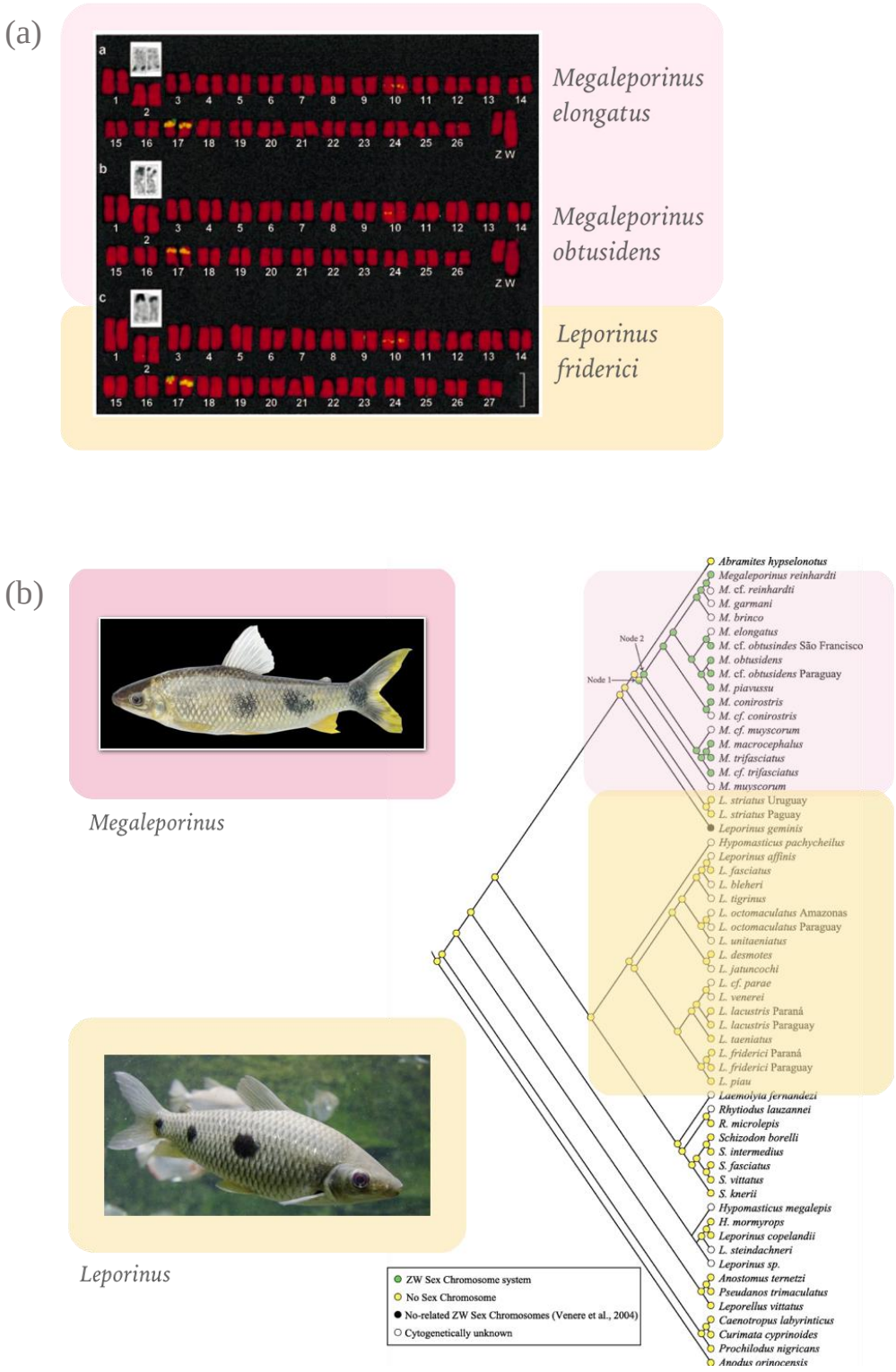


Figura 6. Comparação entre os gêneros *Megaleporinus* e *Leporinus* em especial quanto ao seu cariótipo e cromossomos sexuais apesar da proximidade filogenética. Cariótipos (a) de *M. elongatus, M. obtusidens* (fundo vermelho) e *L. friderici* (fundo amarelo). Note a semelhança entre seus conjuntos cromossômicos, com exceção dos pares sexuais ZW em *Megaleporinus* e seus cromossomos W gigantes. Fonte: adaptado de Martins and Galetti (1999) Filogenia dos dois gêneros (b), mostrando sua proximidade filogenética e respectivas ocorrências de sistemas sexuais. Fonte: adaptado de Ramirez *et al.* (2017a).

A espécie *Megaleporinus elongatus* é ainda um caso além, uma vez que (Parise-Maltempi *et al.*, 2007) levantaram a suspeita da existência de um sistema múltiplo de cromossomos sexuais do tipo $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$. Assim, foi sugerido que seus cromossomos sexuais estariam em um outro processo de diferenciação comparado a seus congêneres. Apesar de compartilhar um ancestral comum com o restante do gênero e com *Leporinus* (Ramirez *et al.*, 2017b), *M. elongatus* possui características moleculares distintas na composição de seu cromossomo heteromórfico que levariam à diferenciação de seus sexuais em um sistema múltiplo (Hashimoto *et al.*, 2009; Parise-Maltempi *et al.*, 2013; Crepaldi and Parise-Maltempi, 2020).

A presença dos cromossomos sexuais heteromórficos em *Megaleporinus* e, por consequência, sua ausência em *Leporinus* (Figura 6) tornaram estes grupos ótimos exemplos para estudos da evolução e diferenciação de sistemas de cromossomos sexuais, tidos como relativamente recentes em termos evolutivos (Parise-Maltempi *et al.*, 2013; Ramirez *et al.*, 2017b; Crepaldi and Parise-Maltempi, 2020).

É aceita a hipótese do acúmulo massivo de elementos repetitivos nos cromossomos W como sendo um dos maiores responsáveis pela sua diferenciação em *Megaleporinus* (Parise-Maltempi *et al.*, 2007; Da Silva *et al.*, 2012; Marreta *et al.*, 2012; Dulz *et al.*, 2024). O acúmulo de algumas sequências pontuais e a comparação por sondas de cromossomos totais entre as espécies iniciaram a busca por este conteúdo, inclusive demonstrando diferenças em cada cromossomo W e denotando mais um grau de complexidade no que concerne a formação dos sexuais dentro do gênero (Nakayama *et al.*, 1994; Parise-Maltempi *et al.*, 2007; Hashimoto *et al.*, 2009; Utsunomia *et al.*, 2019; Crepaldi and Parise-Maltempi, 2020; Crepaldi *et al.*, 2021).

Apesar da semelhança morfológica dos cromossomos W de *Megaleporinus* e da origem comum atribuída a estes cromossomos (Figura 6), cada espécie possui um sexual com composição e distribuição repetitiva específica (Marreta *et al.*, 2012; Crepaldi and Parise-Maltempi, 2020; Crepaldi *et al.*, 2021; Dulz *et al.*, 2024). Utilizando ferramentas bioinformáticas e técnicas refinadas de prospecção de sequências repetitivas, foi possível acessar o conteúdo de DNAs satélites (satDNAs) de espécies de *Megaleporinus* maneira mais abrangente, bem como avaliar quais as implicações nos seus sistemas sexuais ao buscar detalhes da composição dos mesmos (Utsunomia *et al.*, 2019; Crepaldi and Parise-Maltempi, 2020; Crepaldi *et al.*, 2021; Dulz *et al.*, 2024).

Assim, os gêneros *Megaleporinus* e *Leporinus* podem trazer importantes contribuições para o entendimento da estrutura genômica e de elementos repetitivos na diferenciação de

sexuais em peixes, e a presente tese de doutorado pretende dar continuidade a trabalhos em nosso laboratório sobre este grupo de peixes e seus cromossomos sexuais. No pretende trabalho, investigamos a ocorrência e a distribuição de sequências de DNAs repetitivos, como satDNAs e TEs, em genomas de peixes destes dois gêneros. Também buscamos relacionar estas sequências ao surgimento e diferenciação de sistemas sexuais dentro deste grupo, a fim de entender as características genômicas destes elementos repetitivos que venham a estar relacionadas com sistemas de cromossomos sexuais diferenciados.

2 OBJETIVOS

Os objetivos e metas gerais do trabalho foram:

(i) Investigar a ocorrência e a história evolutiva de DNAs repetitivos (satDNAs e TEs) e sua possível influência na diferenciação em cromossomos sexuais em três espécies de peixes neotropicais com sistemas sexuais distintos: *Megaleporinus elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$) e *Megaleporinus macrocephalus* (ZZ/ZW) (ambos com cromossomos W altamente diferenciados); e *Leporinus friderici* (espécie de gênero próximo, sem cromossomos sexuais heteromórficos).

(ii) Examinar os conteúdos repetitivos mediante um estudo comparativo dessas composições (anotação e análise do conteúdo total de satDNAs e transposons nos genomas) obtidos a partir de sequenciamento genômico Illumina, utilizando abordagens citogenômicas com ferramentas bioinformáticas. A partir dessa composição total, fazer inferências quanto a relação de sequências específicas nos cromossomos sexuais através de mapeamento cromossômico por FISH e comparações genômicas.

(iii) Avaliar a influência dessas sequências repetitivas nos sistemas de cromossomos sexuais e na possível diferenciação dos cromossomos heteromórficos de *Megaleporinus.*

Objetivos específicos

- Descrever o conteúdo de satDNAs de *Megaleporinus elongatus* de modo a avaliar sua riqueza de sequências, sua localização cromossômica e a maior representatividade desse conteúdo nas fêmeas. Compreender a dispersão e acúmulo dessas sequências nos cromossomos entre os sexos buscando avaliar se existe localização preferencial no cromossomo sexual heteromórfico.

- Investigar a história evolutiva dos satDNAs mais abundantes e relevantes à diferenciação dos cromossomos heteromórficos de *M. elongatus* em comparação com macho e fêmea da espécie *Megaleporinus macrocephalus* (ZZ/ZW), integrando

resultados já conhecidos para ambas as espécies e discutindo sua dinâmica nos cromossomos heteromórficos das duas fêmeas;

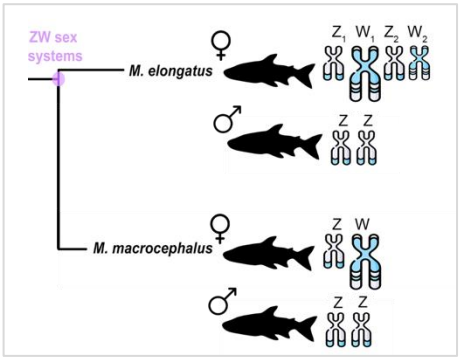
- Descrever o conteúdo de TEs em fêmeas de *M. elongatus, M. macrocephalus* e indivíduo não-sexado de *L. friderici*, a fim de comparar a diversidade e a abundância de TEs nos genomas das três espécies. Buscar entender a dinâmica temporal de dispersão e acúmulo nos três genomas, avaliando histórias evolutivas e se acompanham as filogenias prévias descritas para as espécies e para o surgimento dos cromossomos sexuais.

3 INTRODUÇÃO AOS CAPÍTULOS

Os resultados obtidos levaram a organização de dois trabalhos científicos que serão apresentados nos Capítulos 1 e 2. Ao final de cada capítulo, os links para os respectivos materiais suplementares estão disponíveis para acesso online.

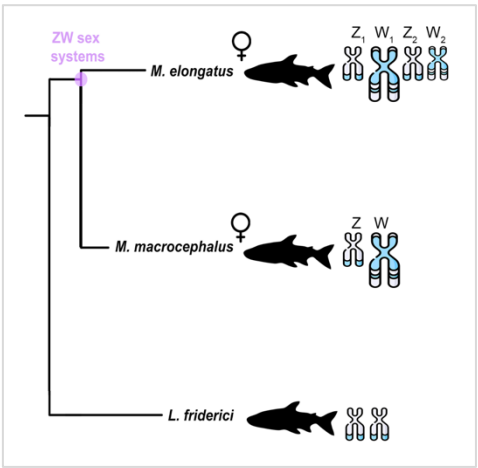
O Capítulo 1, intitulado "**Genomic differences between the sexes in a fish species seen through satellite DNAs**", foi publicado na revista científica Frontiers in Genetics (ISSN 1664-8021) em 2021 (Doi: 10.3389/fgene.2021.728670). Este trabalho descreve os resultados da pesquisa com satelitoma de *Megaleporinus elongatus*, a comparação destes satélites entre os sexos e a análise mais aprofundada dos satélites mais abundantes em comparação com *Megaleporinus macrocephalus*.

Capítulo I



O Capítulo 2, intitulado "**Comparative Analysis of Transposable Elements Dynamics in Fish with Different Sex Chromosome Systems**", foi aceito para publicação na revista científica Genome (ISSN 1480-3321) no começo de 2024 (ID de submissão: Manuscript ID: gen-2023-0134). Esse trabalho apresenta os dados obtidos da descrição de elementos transponíveis em *M. elongatus*, *M. macrocephalus* e *L. friderici*, comparando a dinâmica evolutiva desses elementos nas três espécies.

Capítulo II



3.1 CAPÍTULO I: Genomic Differences Between the Sexes in a Fish Species Seen Through Satellite DNAs

Carolina Crepaldi, Emiliano Martí, Évelin Mariani Gonçalves, Dardo A. Martí e Patricia Pasquali Parise-Maltempi

Artigo publicado na revista Frontiers in Genetics, v.12, p. 728670, 2021.

Doi: 10.3389/fgene.2021.728670

Resumo em Português:

Peixes neotropicais possuem características cariotípicas e genômicas altamente diversificadas, apresentando diversos sistemas de cromossomos sexuais com graus variados de diferenciação. No entanto, o conhecimento sobre composição e evolução específicas do sexo ainda é limitado. As DNA satélites (satDNAs) são sequências repetidas em tandem com distribuição genômica abrangente e caminhos evolutivos distintos. Investigar o conteúdo de satDNA pode esclarecer como a arquitetura genômica é organizada em peixes e em seus cromossomos sexuais. O presente estudo investigou o satelitoma de *Megaleporinus elongatus*, um peixe de água doce com um sistema putativo de cromossomos sexuais múltiplos $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$, que inclui um cromossomo W_1 altamente heterocromático e diferenciado. O satelitoma da espécie é composto por 140 famílias diferentes de satDNA, incluindo sequências isoladas anteriormente e novas famílias encontradas neste estudo. Essa diversidade é notável considerando a proporção relativamente baixa que as satDNAs representam no genoma de *M. elongatus* (aproximadamente apenas 5%). Diferenças entre os sexos em relação ao conteúdo de satDNA também foram evidenciadas, sendo essas sequências 14% mais abundantes no genoma da fêmea. A ocorrência de homogeneização dos satDNAs mais abundantes na espécie está intimamente ligada ao enriquecimento destes satélites associados ao W_1 em fêmeas. Embora ambos os sexos compartilhem praticamente todas as satDNAs, a amplificação massiva de apenas algumas delas acompanhou a diferenciação do W_1. Investigamos também a expansão e diversificação das duas satDNAs mais abundantes de *M. elongatus*, MelSat01-36 e MelSat02-26, ambas sequências altamente amplificadas em W_1 e, no caso de MelSat02-26, também presentes nos cromossomos Z_2 e W_2. Comparamos suas ocorrências em *M. elongatus* e na espécie-irmã *M. macrocephalus* (com um sistema padrão de cromossomos sexuais ZW) e concluímos que ambas as satDNAs levaram à formação de conjuntos altamente amplificados em ambas as espécies; no entanto, elas formaram organizações específicas da espécie nos cromossomos sexuais restritos às fêmeas. Nossos resultados mostram como a composição de satDNA é altamente diversificada em *M. elongatus*, contribuindo significativamente para a diferenciação do W_1 e não para a diversidade de satDNA necessariamente. Além disso, o comportamento evolutivo dessas repetições pode estar associado à plasticidade genômica e à variabilidade de satDNA entre os sexos e entre espécies intimamente relacionadas, influenciando como cromossomos sexuais heteromórficos aparentemente homólogos passam por evolução independente de satDNAs.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 September 2021
doi: 10.3389/fgene.2021.728670

Genomic Differences Between the Sexes in a Fish Species Seen Through Satellite DNAs

*Carolina Crepaldi[1], Emiliano Martí[1], Évelin Mariani Gonçalves[1], Dardo Andrea Martí[2] and Patricia Pasquali Parise-Maltempi[1]**

[1]*Departamento de Biologia Geral e Aplicada, Instituto de Biociências (IB), Universidade Estadual Paulista (UNESP), Rio Claro, Brazil,* [2]*Laboratorio de Genética Evolutiva, Instituto de Biología Subtropical (IBS), Universidad Nacional de Misiones (UNaM), CONICET, Posadas, Argentina*

Neotropical fishes have highly diversified karyotypic and genomic characteristics and present many diverse sex chromosome systems, with various degrees of sex chromosome differentiation. Knowledge on their sex-specific composition and evolution, however, is still limited. Satellite DNAs (satDNAs) are tandemly repeated sequences with pervasive genomic distribution and distinctive evolutionary pathways, and investigating satDNA content might shed light into how genome architecture is organized in fishes and in their sex chromosomes. The present study investigated the satellitome of *Megaleporinus elongatus*, a freshwater fish with a proposed $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ multiple sex chromosome system that encompasses a highly heterochromatic and differentiated W_1 chromosome. The species satellitome comprises of 140 different satDNA families, including previously isolated sequences and new families found in this study. This diversity is remarkable considering the relatively low proportion that satDNAs generally account for the *M. elongatus* genome (around only 5%). Differences between the sexes in regards of satDNA content were also evidenced, as these sequences are 14% more abundant in the female genome. The occurrence of sex-biased signatures of satDNA evolution in the species is tightly linked to satellite enrichment associated with W_1 in females. Although both sexes share practically all satDNAs, the overall massive amplification of only a few of them accompanied the W_1 differentiation. We also investigated the expansion and diversification of the two most abundant satDNAs of *M. elongatus,* MelSat01-36 and MelSat02-26, both highly amplified sequences in W_1 and, in MelSat02-26's case, also harbored by Z_2 and W_2 chromosomes. We compared their occurrences in *M. elongatus* and the sister species *M. macrocephalus* (with a standard ZW sex chromosome system) and concluded that both satDNAs have led to the formation of highly amplified arrays in both species; however, they formed species-specific organization on female-restricted sex chromosomes. Our results show how satDNA composition is highly diversified in *M. elongatus*, in which their accumulation is significantly contributing to W_1 differentiation and not satDNA diversity per se. Also, the evolutionary behavior of these repeats may be associated with genome plasticity and satDNA variability between the sexes and between closely related species, influencing how seemingly homeologous heteromorphic sex chromosomes undergo independent satDNA evolution.

Keywords: satellitome, concerted evolution, satDNA evolution, neotropical fish, fish sex chromosomes, megaleporinus, anostomidae

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
Pablo Mora Ruiz, University of South Bohemia in České Budějovice, Czechia
Alexandr Sember, Academy of Sciences of the Czech Republic (ASCR), Czechia
Sebastián Pita, Universidad de la República, Uruguay

****Correspondence:***
Patricia Pasquali Parise-Maltempi
patricia.parise@unesp.br

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 June 2021*
Accepted: *13 September 2021*
Published: *30 September 2021*

Citation:
Crepaldi C, Martí E, Gonçalves ÉM, Martí DA and Parise-Maltempi PP (2021) Genomic Differences Between the Sexes in a Fish Species Seen Through Satellite DNAs. Front. Genet. 12:728670. doi: 10.3389/fgene.2021.728670

INTRODUCTION

Among the repetitive fraction of a eukaryotic genome, satellite DNAs (satDNAs) are one of the most abundant elements, characterized as tandemly organized sequences that can be amplified into multiple copies in the genome (Charlesworth et al., 1994; Plohl et al., 2012). Selective forces act loosely on satDNAs, as they are prone to accumulate random mutations and rapidly diverge from each other. This ultimately leads to large arrays of diversified satellite composition, as they can vary in sequence length, nucleotide composition, genomic position and chromosome distribution (Ugarković and Plohl, 2002; Garrido-Ramos, 2017). Repeats within the same satDNA family, however, show lower divergence rates as they do not evolve independently, but via what is called "concerted evolution" (Dover, 1982; Elder and Turner, 1995; Thakur et al., 2021). These satDNAs are submitted to the process of molecular drive, which causes sequence homogenization for species-specific mutations and results in repetitions evolving in concert with each other (Dover, 1982; Dover, 1986; Elder and Turner, 1995; Ugarković and Plohl, 2002; Thakur et al., 2021). The rates at which these sequences expanded, homogenized and eventually fixed in the genome vary for each satDNA family. These levels of variation depend on a number of factors, such as mutation rate, array size and structure, chromosomal structure and recombination rates (Ohta and Dover, 1984; Elder and Turner, 1995; Plohl et al., 2012). SatDNA evolution encompasses the duality of combining stable homogeneous arrays fixed in a genome and the high dynamism of rapidly replaceable sequences (i.e., turnover) (Ugarković and Plohl, 2002). These features still place satDNA evolutionary perspective under necessary evaluation. Nonetheless, what is certainly known so far is how the evolutionary mechanisms governing these sequences are different in comparison to other genomic elements (Plohl et al., 2012; Thakur et al., 2021).

Heterochromatin is a poorly understood genomic component, perhaps given the specific nature of its repetitive content (composed mostly of satDNAs) (Charlesworth et al., 1994; Garrido-Ramos, 2017). Sex chromosomes are a good example of genomic entities that can experience heterochromatin expansions and contractions along their evolution, as they can be submitted to rapid diversification after colonization by repetitive sequences and they might gradually differentiate from their homologs, becoming heteromorphic sex chromosomes (Chalopin et al., 2015; Palacios-Gimenez et al., 2015; Wright et al., 2016; Yano et al., 2016; Sember et al., 2018; Charlesworth, 2021; Kratochvíl et al., 2021). The rapid spread of repetitive sequences in a heterogametic (Y or W) sex chromosome may occur during the initial phase of its existence and it may facilitate the expansion of regions with ceased recombination, which further helps XY or ZW counterparts to differentiate from each other (Charlesworth, 1991; Bergero and Charlesworth, 2009; Bachtrog et al., 2011; Schartl et al., 2016).

Teleost fishes compose the most speciose group of vertebrates, which present equally diversified spectrum of sex chromosome occurrences scattered throughout the taxa (Devlin and Nagahama, 2002; Volff, 2005; Godwin and Roberts, 2018; Sember et al., 2021). Teleosts display variable degrees of molecular differentiation in their sex chromosomes, even among closely related species; which contrasts with the more uniform sex systems found in birds and therian mammals (Ellegren, 2011; Schartl et al., 2016; Charlesworth, 2019). Having said that, not much is known about what molecular mechanisms underlie this large compendium of sex chromosome occurrences or the evolutionary dynamics and genomic features of teleosts as a whole.

Megaleporinus is a Neotropical fish genus with a conserved karyotype of 2n = 54 chromosomes, female heterogamety (ZW) and, in *M. elongatus* case, a hypothetical $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ multiple sex chromosome system (Parise-Maltempi et al., 2007; Parise-Maltempi et al., 2013). The heteromorphic sex chromosomes for both systems (W and W_1) are highly heterochromatic and they constitute the largest elements in the female karyotype, which makes the understanding of *Megaleporinus* species genomic traits and sex chromosome systems a coveted approach (Nakayama et al., 1994; Parise-Maltempi et al., 2007; Hashimoto et al., 2009; Ramirez et al., 2017a; Ramirez et al., 2017b; Dulz et al., 2020). Heterochromatin content of the genus has sparked interest in previous works that specifically targeted repetitive DNA occurrence in the heteromorphic sex chromosomes (da Silva et al., 2012; Marreta et al., 2012; de Borba et al., 2013; Parise-Maltempi et al., 2013; Poltronieri et al., 2014). However, only recently was it possible to effectively quantify satDNA content and trace their evolutionary features within *Megaleporinus* (Utsunomia et al., 2019; Crepaldi and Parise-Maltempi, 2020).

Thus, in the present study we used low-coverage genomic DNA data yielded from Illumina paired-end sequencing to assess and further investigate the genomic organization of the highly diversified satellitome in *Megaleporinus elongatus*, firstly presented by Crepaldi and Parise-Maltempi (2020). Cytogenomic and haplotype analyses were also used to further investigate possible sex-specific patterns of these satDNAs and their evolutionary pathways. We complemented our survey with an in-depth analysis of W_1-located satDNAs in both *M. elongatus* and its sister species *M. macrocephalus*, focusing specifically on the two most abundant and quantitatively relevant elements of *M. elongatus* satellitome. By complementing previous satDNA and FISH results (Crepaldi and Parise-Maltempi, 2020) with new additions to satellitome data and haplotype networks, we managed to trace a possible evolutionary pathway for these satDNAs in *M. elongatus* and understand how they contributed to the recent differentiation of the heteromorphic sex chromosomes and their possible role in the multiple sex chromosome differentiation. Altogether, our study aimed to integrate thorough genomic analyses to previously published data for the *Megaleporinus* genus and to provide new information regarding the evolution of repetitive genomic composition in the group, as well as satDNA evolutionary differences between the sexes and closely related species.

MATERIALS AND METHODS

Species Sampling and Chromosome Preparation

Chromosomal preparations and tissue samples from three males and three females from *Megaleporinus elongatus* (Anostomidae)

were analyzed. All samples were already available at the Animal Cytogenetics Laboratory in UNESP Rio Claro, Brazil from previous studies (da Silva et al., 2012; Marreta et al., 2012; Parise-Maltempi et al., 2013; Crepaldi and Parise-Maltempi, 2020). All procedures for sampling, material handling and analysis were authorized and approved by the Animal Ethics Committee (Comitê de Ética no Uso de Animais (CEUA) (protocol number 3524, approval code 09/2017), by the Brazilian Institute of Environment and Renewable Natural Resources (Instituto Brasileiro do Meio Ambiente e Recursos Naturais Renováveis - IBAMA) (19833-1) and the Brazilian College of Animal Experimentation (Colégio Brasileiro de Experimentação Animal - 016/04-CEEA). Mitotic chromosomes were obtained from anterior kidney cell suspensions, according to Foresti and Toledo (1981).

Genome Sequencing and Computational Satellitome Analysis

In the present study, previously sequenced libraries from each sex of *M. elongatus* using Illumina® Hiseq™ 2000 platform (female) and Illumina® Hiseq™ 4000 platform (male) (Crepaldi and Parise-Maltempi, 2020) were used, which provided 1.9 Gb of sequence data and yielded 19,289,312 paired-end trimmed reads for the female, and 1.8 Gb of data and 17,837,098 paired-end trimmed reads for the male library.

The sequenced libraries were used for comparative analysis regarding their satDNA content. We focused both on the species' satellitome as a whole and on the differences between the sexes, in search for sequences that could be more representative in the female and probably enriched in the heteromorphic W_1 sex chromosome. To perform a high-throughput analysis, the satMiner bioinformatic protocol for satDNA prospection in both libraries was used (Ruiz-Ruano et al., 2016) available at GitHub (https://github.com/fjruizruano/satminer). The satMiner protocol uses several rounds of clustering in RepeatExplorer (Novák et al., 2013) to identify and extract satDNA sequences, and each round includes filtering out reads matching previously assembled contigs with deconseq 0.4.3 (Schmieder and Edwards, 2011), in order to identify and extract as many sequences as possible, even with low abundance in the genome. We then started with a library sampling of 200,000 reads, incrementing this number by two in each consequent round of RepeatExplorer clustering. For each round, we selected clusters with spherical shaped graphs for putative satDNA. Each selected cluster was manually analyzed for their internal contig structure and tandem repetitions were investigated using the dotplot tool implemented in Geneious v4.8 (Drummond et al., 2015) and Tandem Repeat Finder (TRF) (Benson, 1999).

The clustering and filtering steps were repeated 13 times for the female library and 9 times for the male one, adding new filtered reads in each iteration until we could no longer detect new satDNAs in neither. A parallel homology search was performed in both male and female rounds using previously detected satDNA consensus sequences by Crepaldi and Parise-Maltempi (2020) as a custom library to match ones that might have previously been isolated and physically mapped.

After satDNA mining, all-against-all alignments with RepeatMasker (Smit et al., 2013) were performed to search for homologous satDNAs, and by comparing all monomers from all clusters we were able to classify them into superfamilies, families or variants (Ruiz-Ruano et al., 2016). Non-redundant consensus library was used for each satDNA family to check for possible similarities with published sequences deposited in Genbank and Repbase employing BLAST (http://www.ncbi.nlm.gov/Blast/) and Censor (http://www.girinst.org/censor) searches.

All satDNA families were numbered in order of decreasing abundance in the female genome and assigned following the nomenclature proposed by Ruiz-Ruano et al. (2016). All satDNAs that were previously isolated by Crepaldi and Parise-Maltempi (2020) were properly renamed also following this criterion. Sequences are deposited in GenBank under accession numbers MZ546645–MZ546784.

RepeatMasker with rmblast engine was used to determine abundance and average nucleotide divergence (Kimura-2-parameter, K2P) for each variant in both sexes. We estimated the genomic abundance for every satDNA in the male and female libraries as the number of nucleotides aligned to the reference consensus divided by the library size (in bp). With this data we generated repeat landscapes for the relative abundance (Y-axis) at 1% intervals of K2P distance from the consensus (X-axis), using the script calcDivergencFromAlign.pl (from RepeatMasker utils). A subtractive landscape was subsequently generated to evaluate which satDNA families differ between both libraries, in turn providing the first indications of which satDNA are more prominent in one sex in comparison to the other.

Physical Mapping via FISH in Male Metaphases

We selected the same previously isolated and amplified 52 satDNAs in the female (Crepaldi and Parise-Maltempi, 2020) and amplified them via PCR in *M. elongatus* males following the protocol described by the authors. The PCR products were confirmed via Sanger sequencing.

The sequences of each satDNA obtained through PCR were labeled by nick translation with digoxigenin-11-dUTP (Roche®) or biotin-14-dATP (Invitrogen®). Fluorescence *in situ* hybridization (FISH) experiments were performed following the method described by Pinkel et al. (1986), with small adjustments described by Crepaldi and Parise-Maltempi (2020). The resulting slides were visualized under an Olympus® BX51 fluorescence microscope, with a digital camera Olympus® DP71 attached, and the images were captured using the DP Controller camera software. For each slide, a minimum of 20 metaphases were analyzed and photographed to confirm the FISH results.

Sex-Biased Ratio

The different enrichment of all satDNAs across the sexes was determined by generating a female to male ratio as we calculated the quotient between the abundance values of each satDNA family. This data complemented the subtractive landscape by providing more between-sexes differences, as satDNA families

with Female/Male (F/M) ratio higher than one were considered more abundant in females (as the threshold to determine it as more prevalent in this sex).

RepeatProfiler pipeline (Negm et al., 2021) was applied to generate comparative variant-enhanced profiles of selected satDNAs, which provide a summary of variant sites that are relative to the consensus sequences and may uncover sex-specific signatures of sequence variants and possible point mutations in satDNAs of interest. The 22 satDNA families that are most enriched in each sex considering the F/M ratio were selected for profiling. RepeatProfiler relies on Bowtie2 (Langmead and Salzberg, 2012) for mapping the reads to the consensus monomers of the selected sequences and the pipeline generated a PDF file for each selected satellite with the variant-enhanced profiles for both sexes as an output. We applied Bowtie2 default parameters for RepeatProfiler.

After analyzing the resulting profiles, we generated individual landscapes for each selected female-biased satDNAs to confirm different amplification and divergence of their copies in male and female genomes.

Retrieving satDNA Monomers From Raw Reads of *M. elongatus* and *M. macrocephalus*

Comparative cytogenetic analysis of the two most abundant satDNAs (MelSat01-36 and MelSat02-26) revealed particular characteristics that prompted us to investigate them further, such as differences in clustered patterns in the W chromosome in *M. macrocephalus* and W_1 in *M. elongatus* and occurrence in autosomal pairs in both males and females. MelSat02-26 is specifically interesting given it is a fragment of Le*Spe*I, a repetitive marker used for the *M. elongatus* hypothetical multiple sex chromosome system (Parise-Maltempi et al., 2007; Crepaldi and Parise-Maltempi, 2020).

Monomers from Illumina reads representing MelSat01-36 and MelSat02-26 were extracted from both sexes of *M. elongatus* and *M. macrocephalus* libraries. Thus, Minimum Spanning Trees (MSTs) of these satDNA families were generated to trace the diversification patterns of copies between sexes and species. First, we mapped reads from raw Illumina libraries from the two selected satDNA, of both sexes of *M. elongatus* and *M. macrocephalus*, with a custom script (https://github.com/fjruizruano/satminer/blob/master/mapping_blat_gs.py).

Libraries of *M. macrocephalus* were retrieved from Sequence Read Archive (SRA) under the accession numbers SRR7263033 and SRR7263034. Mapped reads were then extracted from SAM files and aligned separately using MUSCLE (Edgar, 2004) with default parameters. The resulting alignment files of each satDNA were used as input in PHYLOViZ 2.0 (Nascimento et al., 2017) to construct MSTs on the bases of pairwise differences. RepeatMasker was used to determine abundance and average nucleotide divergence (K2P) for MelSat01-36 and MelSat02-26 in both sexes of *M. macrocephalus*. Then, the genomic abundance of these two satDNAs in reference to this species library sizes (in bp) was calculated, and this data was used to generate comparative landscapes between *M. elongatus* and *M. macrocephalus*.

RESULTS

General Satellitome Analysis in *M. elongatus*

A total of 140 different satDNA families (308 variants) were uncovered for *M. elongatus* as a whole, with a predominance of short repeat unit lengths (RUL) ranging from 11 to 245 bp (average of 43 bp) for both sexes. A + T content of consensus sequences varied between 30.8 and 80.4% (60% on average), which indicated a slight tendency towards A + T rich content for the majority of satDNAs (**Supplementary Table S1**). The homology analysis of repeat units revealed only one satDNA superfamily (SF1) comprising MelSat09-60 and MelSat113-60 (76.7%) (**Supplementary Figure S1**) and present in both sexes with relatively similar abundances. Both sequences share the same RUL (60 bp) and have high divergence values in both sexes (**Supplementary Table S1**).

SatDNA genome proportion ranged from 0.0001 to 0.484%, with the three most abundant satDNAs (MelSat01-36, MelSat02-26 and MelSat03-177) showing an abundance higher than 0.3%. SatDNA abundance in the present work regards the values determined by RepeatMasker that applied a substantial amount of reads in comparison to the pool of randomly selected reads analysed previously (and solely via RepeatExplorer output) by Crepaldi and Parise-Maltempi (2020). MelSat01-36 and MelSat02-26 remained the two most abundant families as previously described (Crepaldi and Parise-Maltempi, 2020) (0.484 and 0.413% respectively), but MelSat03-177 (previously the family MelSat07-177 located in the centromeric area) ranked third in abundance after this thorough analysis, representing 0.338% of the genome.

Therefore, the satellitome of *M. elongatus* consists primarily of the top three most prevalent satDNA, which that comprise almost 25% of the whole compendium of satellites, in addition to 10 families with abundance between 0.1 and 0.3% and a remainder of rare or very low abundant sequences (**Supplementary Table S1**). Overall divergence values were relatively variable for the species as a whole, ranging from 2.85 to 33.02% (average divergence for the species was 10.73%).

Searches in GenBank resulted in 46 families from *M. elongatus* with positive results with deposited *M. macrocephalus* satDNA sequences (Utsunomia et al., 2019). BLAST results and subsequent alignments for MelSat03-177 showed high similarities with centromeric sequences from other Characiformes (Utsunomia et al., 2017, Utsunomia et al.2019). MelSat40-52 (previously named MelSat49-52 (Crepaldi and Parise-Maltempi, 2020) also showed positive results for other satellite sequences in Characiformes, a conserved sequence with active transcription in *Characidium gomesi* (dos Santos et al., 2021).

FISH analysis was performed in male metaphases for *M. elongatus* for the same satDNA previously mapped in the female (Crepaldi and Parise-Maltempi, 2020). From the 52

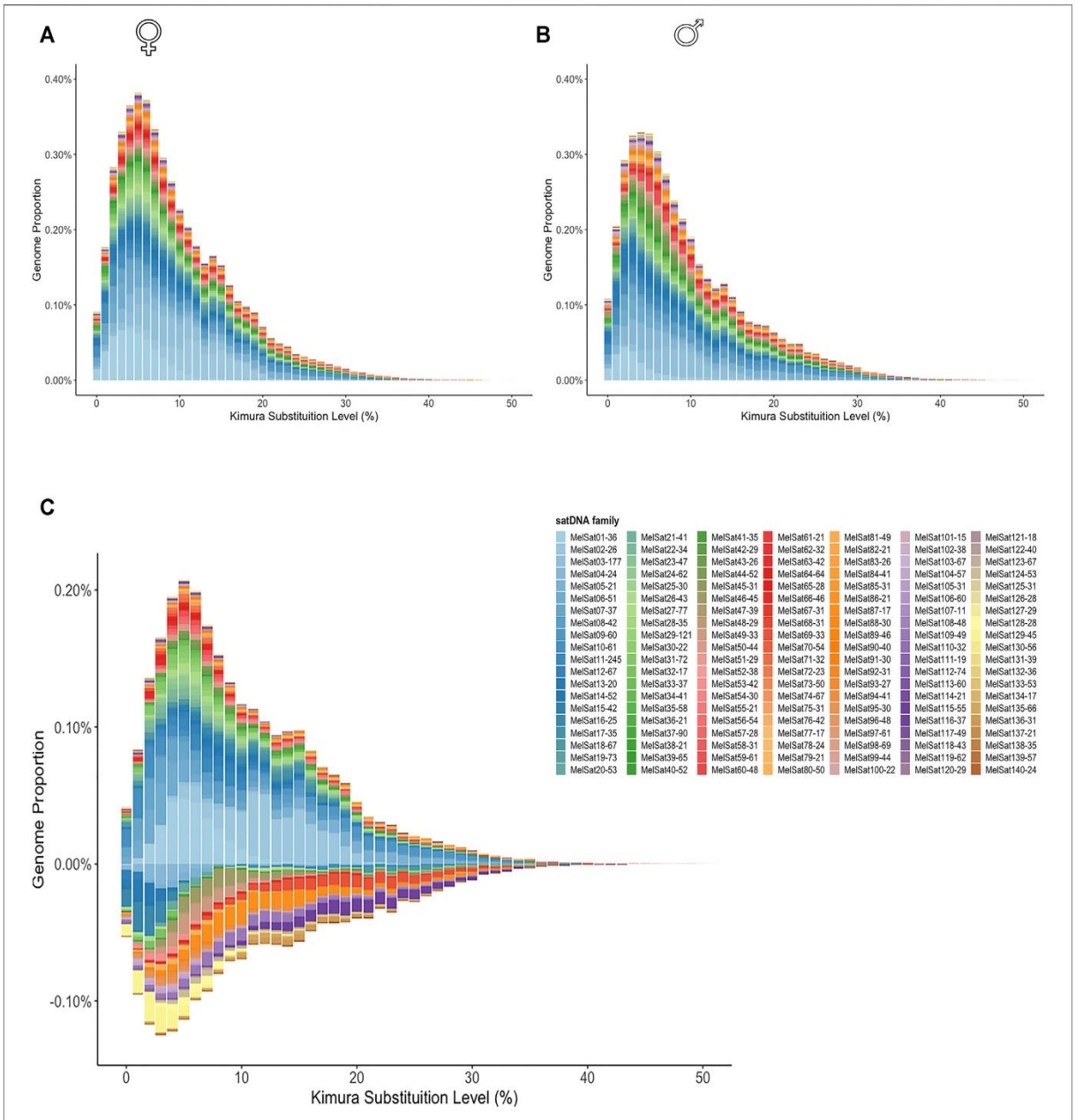


FIGURE 1 | Repeat landscapes (abundance vs. divergence) for satDNAs identified in female **(A)** and male **(B)** *M. elongatus*. The graphs show, for each color-coded element, the sequence divergence according to Kimura distance (x-axis) in relation to their copies in the genome (y-axis). Copies clustered to the left (lower divergence) potentially correspond to recent copies occurring in the genome. A subtractive landscape **(C)** was obtained by subtracting male library (negative values) from the female one (positive values). Notice the diversity of satDNA families also present in the male genome in the subtractive landscape.

satDNAs positively amplified via PCR, only 5 (MelSart02-26, MelSat22-34, MelSat29-121, MelSat44-52 and MelSat52-38) successfully hybridized in this sex (**Supplementary Figure S2**), in one autosomal pair each and, for MelSat02-26, in Z_2Z_2. Each satDNA presented a single band in the chromosomes, and positioned either in the telomeric or in the centromeric regions. No satDNA was mapped exclusively in the male, as all five presented FISH bands in female chromosomes (Crepaldi and Parise-Maltempi, 2020).

Satellitome Differences Between the Sexes

The total satDNA composition corresponded to 4.83 and 4.23% of the female and male genomes, respectively. Average satDNA family divergence was lower in the female (10.38%) than in the

male genome (11.07%), and some satDNAs had quite higher divergence values for the male in comparison to the female, such as MelSat66-46 (33.02 and 7.38%, for male and female respectively) and MelSat64-64 (22.44 and 6.34%). It was later found these satDNAs are highly female-biased according to the F/M ratio (**Supplementary Table S1**) and with insignificant abundance values for the male library, which sparked our interest to investigate further these differences, described hereinafter.

We generated individual repeat landscapes for female and male *M. elongatus* (**Figure 1A** and **Figure 1B**, respectively) and a subtractive landscape (**Figure 1C**). The subtractive landscape revealed higher proportions of several abundant satDNA families in the female library, such as MelSat01-36 and MelSat02-26, which are located in the W1 sex chromosome (Crepaldi and Parise-Maltempi, 2020). Interestingly, we could also confirm that the male library presents a higher variety of male-biased satDNAs in comparison to the female (**Figure 1C**), which are located in the W_1 sex chromosome (Crepaldi and Parise-Maltempi, 2020). Interestingly, also through the subtractive repeat landscape we could confirm that the male library presents a higher variety of male-biased satDNAs in comparison to the female (**Figure 1C**).

The computational analysis revealed that, except for MelSat131-39, which is present in the female genome only, all remaining 139 satDNA families are shared between the sexes. From these, however, 124 are differently enriched across sexes as 45 satDNAs had a F/M ratio higher than 1, suggesting an enrichment in the female library, while 79 were deemed male-biased (F/M ratio lower than 1). Despite having less overall female-biased satDNAs in numbers, several families had ratios ranging from 13 up to 347 in the female, denoting an expressive enrichment in this sex compared to the male library (**Supplementary Table S2**). In short, the female library has the most abundant satDNAs, enriched in the W_1 (Crepaldi and Parise-Maltempi, 2020); however, the greater diversity of satDNAs is male-biased, comprised mostly of the very rare and less abundant families of the satellitome.

Variant Profiles for Sex-Biased satDNAs

RepeatProfiler pipeline was applied in order to evaluate the sequences of the most sex-biased satDNAs and how they might differ between the sexes. The output provided variant-enhanced profiles for each selected satDNA, summarizing sex-specific signatures for some sequences (**Supplementary Figures S3, S4**). All male-biased satDNAs (**Supplementary Figure S3**) have relatively similar divergent values for both sexes and, with few exceptions (such as MelSat140-24, MelSat137-21 and MelSat92-31), they did not present such apparent differences between sexes as the female-biased ones, showing some conservatism in comparison with male and female profiles. Female-biased satDNAs (**Supplementary Figure S4**), on the other hand, showed some reoccurring patterns: 1) most have discrepant divergent values between the sexes (with high divergent values for the male) and are either absent in males or have incomplete profiles for this sex, since the aligner failed to map high-divergent reads to consensus; 2) comparable satDNAs were prone to variation in male variants, with particular positions exhibiting almost fixed copies, since mutation profiles differed between the sexes when abundance values dropped; and 3) the only two satDNAs that presented relative homogeneity between the sexes are MelSat01-36 and MelSat02-26 (the two most abundant satDNAs in the genome).

We complemented the variant profiles results by generating individual repeat landscapes (**Supplementary Figure S5**) for the top most female-biased satDNAs, to check for differential amplification and divergence values between the sequences especially on satDNAs that are most likely specific to the female. The landscapes also evidenced the higher abundance in the female and contrasting divergences for these sequences in the male genome. Some satDNAs, such as MelSat57-28, MelSat112-74 and MelSat123-67 showed monomers with low divergences in the female and shared higher divergences between the sexes, indicating highly divergent copies comparing the genomes.

Tracing the Amplification of W-Localized satDNAs in *M. elongatus* and *M. macrocephalus*

MSTs for MelSat01-36 and MelSat02-26 were generated and complemented with comparative landscapes for these sequences individually in order to integrate their abundance and divergence values to the analysis. Different MSTs and line plots were obtained for each satDNA (**Figure 2**).

The topology for MelSat01-36 confirms the relatedness and conservatism of this satDNA, especially when considering the larger haplotype shared between both species and its co-localized pattern in FISH results (**Figure 2A**). Furthermore, due to its low divergence values in both species, it most likely had a common origin in both genomes and an equally recent amplification in the heteromorphic sex chromosomes in the females.

Some discrepancies, however, are notable for *M. elongatus*. Firstly, clusterisation is absent in the male (no FISH signals). In addition to the fact that it has the least abundance in the male of *M. elongatus* of all four analyzed individuals, it is probably gradually losing copies in its genome. This coincides with another *M. elongatus* conspicuous characteristic, perhaps the most noticeable: the female copies. MelSat01-36 is not only more abundantly represented in the female genome, but it is also presented in two separate clusters in the heteromorphic chromosome (**Figure 2A**).

For MelSat02-26, the resulting MSTs displayed a more divergent pattern and a different evolutionary scenario for this satDNA (**Figure 2B**) was observed. Firstly, no monomers were retrieved from the male *M. macrocephalus*, as its abundance was virtually non-existent in this individual. For the remaining copies, all three (female *M. macrocephalus,* male and female *M. elongatus*) presented similar divergence with rather higher values, which implies that MelSat02-26 comprised an older satDNA especially in comparison to MelSat01-36.

Both sexes of *M. elongatus* share autosomal clusters and its origins appear concomitant in male and female, especially

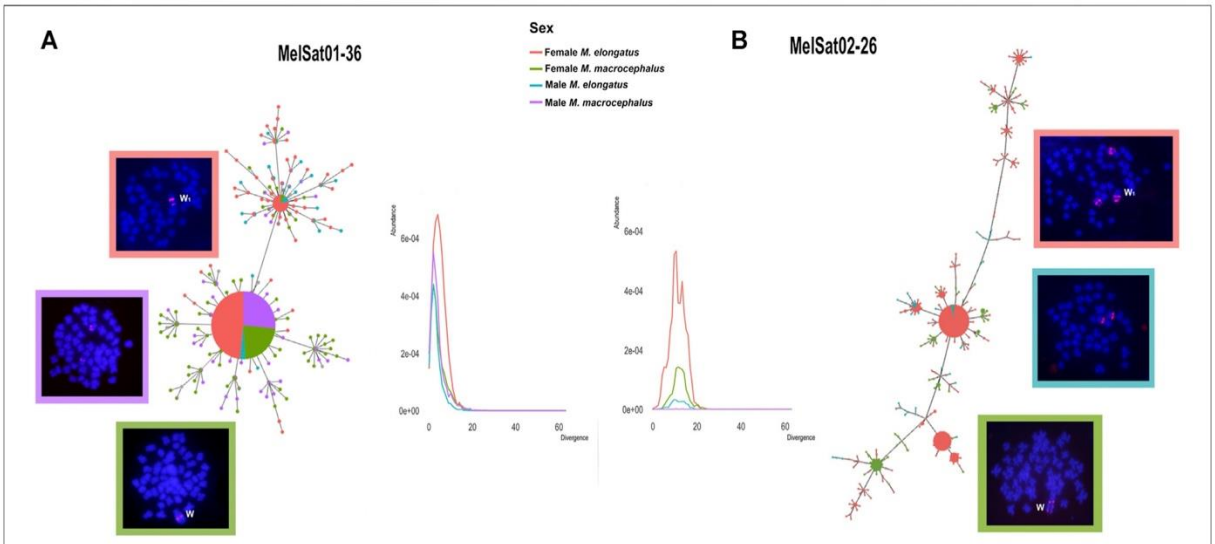


FIGURE 2 | MSTs showing the relationships between isolated monomers for MelSat01-36 **(A)** and MelSat02-26 **(B)** retrieved from *M. elongatus* and *M. macrocephalus* males and females. Respective FISH results and the heteromorphic sex chromosomes W and W_1 are indicated. Male FISH results for *M. elongatus* and *M. macrocephalus* are absent for MelSat01-36 and MelSat02-26 as no FISH signals were obtained, respectively. Abundance x divergence landscapes that shows the respective similar origins in both species are also indicated, but the substantial abundance in *M. elongatus* female for both satDNAs is apparent. Note the different divergences between the satDNAs, as MelSat02-26 has larger divergence values than MelSat01-36 and a rather dispersed organization in comparison to the latter.

considering it is clustered in the same autosomal pair and putative multiple sex chromosomes in both sexes (Parise-Maltempi et al., 2007; Crepaldi and Parise-Maltempi, 2020). There is a distinct spike in abundance for *M. elongatus* female in the line plot for this satDNA (**Figure 2B**), which, paired with the clustered pattern in FISH, shows the substantial amplification in the heteromorphic sex chromosome, which also occurred long since. It presents a different amplification trajectory between the species: although *M. macrocephalus* female also presents clusterisation in its W, it is not as considerable as in *M. elongatus* W_1 and it does not share any copies with *M. macrocephalus* male.

DISCUSSION

The present study provides relevant information regarding the satDNA content of *Megaleporinus elongatus* and its genomic features, adding notable details to the first glimpse of these sequences previously published by our group (Crepaldi and Parise-Maltempi, 2020). Using more thorough methods for satDNA prospection, we expanded the array of satellite features in the species, compared satellitomes between the sexes and uncovered the evolutionary pathway of the most abundant (and most quantitively relevant) satDNAs in the genome and their differences between species.

Firstly, some distinct characteristics are responsible for the *M. elongatus* satellitome as a whole, which showed small and highly diversified sequences, mostly recently amplified (low divergence) and with general low abundance. These characteristics were also found in other Characiformes, such as *Astyanax* (Silva et al., 2017), *Megaleporinus macrocephalus* (Utsunomia et al., 2019) and *Characidium gomesi* (Serrano-Freitas et al., 2020).

The disparity of a large satDNA diversity in *M. elongatus* (140 families) and such a scarce occurrence result in very rare sequences, organized in small arrays and with a dispersed organization in the genome. The massively amplified MelSat01-36 and MelSat02-26 contrasts with the remaining satellitome and indicate that satDNA diversification in *M. elongatus* increases as their clusters become smaller and more dispersed. Fast satDNA turnover may be another possibility, since almost all satDNA content is non-homologous in *M. elongatus*. *M. macrocephalus* presents 17 satDNA superfamilies (Utsunomia et al., 2019), but for *M. elongatus*, on the other hand, we recovered only one superfamily, which also endorses its highly diversified satDNA collection.

Also, most satDNAs were not detectable by physical mapping in neither sex, as presented in our present analyses and previous FISH results (Crepaldi and Parise-Maltempi, 2020). This confirms the majority of the satellitome for *M. elongatus* is arranged in small arrays below the detection threshold of FISH, corroborating other studies (Ruiz-Ruano et al., 2016; Bardella et al., 2020; Ferretti et al., 2020). SatDNA abundance is prone to rapidly change due to molecular mechanisms, such as dispersion and amplification (Garrido-Ramos, 2017), and can result in rapid repatterning as they expand or decrease their arrays (Plohl et al., 2008; Palacios-Gimenez et al., 2020; Sproul et al., 2020; Thakur et al., 2021) and this certainly is the case for *M. elongatus* as well.

This is a rather peculiar satellitome, especially considering our sex-biased analysis. Given the extraordinary heterochromatin content present in the female (owing to the heteromorphic W_1

sex chromosome) (Galetti et al., 1995), it was initially assumed by us that the largest set of satDNAs would be female-biased as previously observed in *M. macrocephalus* (Utsunomia et al., 2019); however, not many satDNAs (in terms of number of different families) accounted for female-biased or W_1-specific sequences in *M. elongatus*.

In what regards the male genome, satDNAs are equally abundant in qualitative terms; however, they are most likely becoming gradually lost and/or constrained instead of actually accumulating. Their low copies in the genome increase the possibility to escape homogenization mechanisms (Elder and Turner, 1995; Ugarković and Plohl, 2002; Plohl et al., 2012) as we noticed different homogenization patterns between the sexes. Repeats in larger arrays (and clustered in the W_1 chromosome) had higher homogenization rates, and low abundance sequences (especially the male-biased ones) presented a more divergent pattern. It is possible to assume some non-exclusive explanations for this scenario. Firstly, more heterogeneity can be expected among repeating units if the mutation rate is high, relative to the rate in which a variant spreads through an array (Lorite et al., 2004; Ruiz-Ruano et al., 2019). Also, the heteromorphic W_1 chromosome has a tendency to accumulate satDNAs already present in the female genome (Molina and Galetti, 2007; Crepaldi and Parise-Maltempi, 2020), which results in a quick, massive amplification in this sex in comparison to the poor clusterisation in the male counterpart (cf. Lower et al., 2018; Vondrak et al., 2020). In conclusion, satellitome diversity does not seem to be the main factor leading to W_1 heterochromatic expansion in *M. elongatus*, instead the high amplification and homogenization rates of few particularly abundant satDNAs in the female genome effectively contributed to heterogametic sex chromosome differentiation.

Our second approach regarded a more evolutionary focus in the sex chromosome system and its satDNA content in *Megaleporinus elongatus,* as sex chromosome dynamics analysed through sex-linked repetitive DNA profiles has been shown to be effective (Traut and Winking, 2001; Schemberger et al., 2011; Cioffi et al., 2012; Terencio et al., 2012; Yano et al., 2016; Yano et al., 2017; Sember et al., 2018; Deakin et al., 2019; Schemberger et al., 2019). The heteromorphic sex chromosomes (W and W_1) in *Megaleporinus* have already been deemed derived from a common ancestral chromosomic pair and differentiated from the Z chromosomes via massive heterochromatinization (Galetti and Foresti, 1986; Marreta et al., 2012; Parise-Maltempi et al., 2013; de Barros et al., 2018). However, variation in repetitive content within the genus could be caused by the expansion and contraction of these sequences (Utsunomia et al., 2019; Crepaldi and Parise-Maltempi, 2020).

Regardless of the library size of satDNAs, usually one or very few satDNA families are the most predominant in each species (Ruiz-Ruano et al., 2016; Silva et al., 2017; Bardella et al., 2020; da Silva et al., 2020) and the best candidates were none other but the two most abundant satDNA in *M. elongatus*, MelSat01-36 and MelSat02-26. Both satDNAs have concomitant origins in both sexes of *M. elongatus* and *M. macrocephalus,* but they have different evolutionary pathways in each species and concerted evolution might be acting separately for each of them.

MelSat01-36 is shared by W and W_1, but it is clustered only in males of one species (with prominent FISH signals being present in *M. macrocephalus* but absent in *M. elongatus*). Its two distinct clusters in female *M. elongatus* prompted us to combine all of our data on this satDNA and attempt to trace its relatedness between the species. All individuals presented similarly low divergence values for this satDNA, indicating a recent amplification for both species. Although identical sequences for MelSat01-36 are shared between the species, copy number can vary dramatically even among related organisms, without necessarily varying in their nucleotide sequences (Ugarković and Plohl, 2002; Louzada et al., 2020). Given our MST for this satDNA (**Figure 2**), ancestral species for *M. elongatus* and *M. macrocephalus* most likely presented the same conformation as *M. macrocephalus,* with one cluster in the ancestral W sex chromosome in the female and another cluster in a male autosomal chromosome. For *M. elongatus*, the male has lower copy numbers and a dispersed organization, since it did not present any FISH signals. This loss of autosomal clusters might have contributed to the homogenization in W_1 in *M. elongatus,* and generated the enriched and duplicated regions in this chromosome, fixating in the female genome as described for many other organisms (Dalíková et al., 2017; Palacios-Gimenez et al., 2017; da Silva et al., 2020; Ferretti et al., 2020; Rovatsos et al., 2021).

MelSat02-26, on the other hand, is an older satDNA, with higher divergence values for both species and it is practically absent in male *M. macrocephalus*. Despite shared between the heteromorphic W and W_1, it has a much less conserved pattern (**Figure 2**) in comparison to MelSat01-36, with dispersed clusters and almost none shared haplotypes between the females. An ancestral library for this satDNA cluster was most likely present in both the autosomes and the heteromorphic chromosome, much like MelSat01-36, however remaining highly amplified on both sexes of *M. elongatus*. It most likely had more time to diverge between the species and fixate in *M. elongatus* as well-defined clusters, as it is also co-localized in the putative multiple sex chromosomes (Z_2 and W_2). What previously seemed like having different accumulation in terms of time for each chromosome (Crepaldi and Parise-Maltempi, 2020), now, with more intimate evolutionary details, we can see that MelSat02-26 seems to have emerged equally in all *M. elongatus* chromosomes; along with its exclusive clusterisation pattern in the female Z_2 and W_2, shared with no other autosomal satDNA in the species and very particular to this sex.

The dynamics of the most abundant satDNAs in *M. elongatus* demonstrates not only their intimate evolution with the highly differentiated heteromorphic W_1 sex chromosome, but also confirms that not only general molecular satDNA evolution is at play; particular characteristics of the species also concomitantly generate diversified satellitomes (Ugarković and Plohl, 2002; Lorite et al., 2004, Lorite et al., 2017; Richard et al., 2008; Palacios-Gimenez et al., 2017; Cabral-de-Mello et al., 2021). The ZW-based multiple sex chromosome system in *M. elongatus* shares a conserved heteromorphic W chromosome with the remaining *Megaleporinus*; but as seen by the two most prevalent satDNAs in the species their interspecific differences show distinct paces for W/W_1 differentiation.

SatDNA evolution seems to be not only recent but very fast-paced in the genus, with high sequence turnover rates, and this might contribute to an equally fast and independent differentiation process in their young sex chromosome systems. A clear outcome of this is the highly amplified MelSat02-26 in Z_2 and W_2 in *M. elongatus* female, an interesting satDNA on its own as it partially represents a bigger repetitive sequence and molecular marker (Le*Spe*I) for the putative multiple chromosomes (Parise-Maltempi et al., 2007; Crepaldi and Parise-Maltempi, 2020). While MelSat02-26 is underrepresented in the female Z_1 chromosome, the unusual female-specific clustered pattern shared by W1 and the novel elements (W_2 and Z_2) might indicate its spread via ectopic recombination, which has most likely helped this satDNA expand into large, homogeneous arrays (Ugarković and Plohl, 2002; Lower et al., 2018).

The rare $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ system of *M. elongatus* is so far only found in one other fish species (*Ancistrus dolichopterus*) (de Oliveira et al., 2008; Favarato et al., 2016), and remains a puzzling occurrence even among multiple sex chromosome systems as whole (Sember et al., 2021). Analyzing other *Megaleporinus* as well as the sister genus *Leporinus* (comprising species without heteromorphic sex chromosomes) will eventually broaden our understanding of satDNA diversity and sex chromosome evolution as a whole in this group.

In the present study, we used cytogenomic approaches for satDNA analysis and visualization of their male and female heterogeneity. The results corroborate the recent spike of highly clustered repetitive content in the female genome and showed it has very distinctive characteristics even in comparison to closely-related species. Also, the recent burst of repetitive satDNA in the multiple sex chromosome system is still an ongoing process under the molecular mechanisms for satDNA evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories and accession number(s) can be found in the article.

ETHICS STATEMENT

The animal study was reviewed and approved by Sao Paulo State University Animal Ethics Committee (protocol number 3524, approval code 09/2017).

AUTHOR CONTRIBUTIONS

CC contributed to the work's conception, data acquisition, analysis and interpretation, and wrote the manuscript. EM contributed to bioinformatic data acquisition, analysis and interpretation. CC and EG conducted cytogenetic analysis. DM contributed to bioinformatic data resources. PPP-M conceptualized, funded and coordinated the research. All authors revised the manuscript.

FUNDING

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #2017/00195_7 and FAPESP #2018/12906-8).

ACKNOWLEDGMENTS

The authors would like to thank Dr. Diogo Cavalcanti Cabral-de-Mello and his laboratory students for the valuable suggestions and constructive comments, and critical reading of an earlier version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.728670/full#supplementary-material

REFERENCES

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W., et al. (2011). Are all sex chromosomes created equal. *Trends Genet.* 27 (9), 350–357. doi:10.1016/j.tig.2011.05.005

Bardella, V. B., Milani, D., and Cabral-de-Mello, D. C. (2020). Analysis of *Holhymenia Histrio* genome provides insight into the satDNA evolution in an insect with holocentric chromosomes. *Chromosome Res.* 28 (3), 369–380. doi:10.1007/s10577-020-09642-1

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi:10.1093/nar/27.2.573

Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24 (2), 94–102. doi:10.1016/j.tree.2008.09.010

Cabral-de-Mello, D. C., Zrzavá, M., Kubíčková, S., Rendón, P., and Marec, F. (2021). The role of satellite DNAs in genome architecture and sex chromosome evolution in Crambidae moths. *Front. Genet.* 12, 661417. doi:10.3389/fgene.2021.661417

Caetano de Barros, L., Piscor, D., Parise-Maltempi, P. P., and Feldberg, E. (2018). Differentiation and Evolution of the W Chromosome in the Fish Species of *Megaleporinus* (Characiformes, Anostomidae). *Sex. Dev.* 12 (4), 204–209. doi:10.1159/000489693

Chalopin, D., Volff, J.-N., Galiana, D., Anderson, J. L., and Schartl, M. (2015). Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Res.* 23 (3), 545–560. doi:10.1007/s10577-015-9490-8

Charlesworth, B., Sniegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371 (6494), 215–220. doi:10.1038/371215a0

Charlesworth, B. (1991). The evolution of sex chromosomes. *Science* 251 (4997), 1030–1033. doi:10.1126/science.1998119

Charlesworth, D. (2021). When and how do sex-linked regions become sex chromosomes. *Evolution* 75 (3), 569–581. doi:10.1111/evo.14196

Charlesworth, D. (2019). Young sex chromosomes in plants and animals. *New Phytol.* 224 (3), 1095–1107. doi:10.1111/nph.16002

Cioffi, M. B., Moreira-Filho, O., Almeida-Toledo, L. F., and Bertollo, L. A. C. (2012). The contrasting role of heterochromatin in the differentiation of sex

chromosomes: an overview from Neotropical fishes. *J. Fish Biol.* 80 (6), 2125–2139. doi:10.1111/j.1095-8649.2012.03272.x

Crepaldi, C., and Parise-Maltempi, P. P. (2020). Heteromorphic sex chromosomes and their DNA content in fish: an insight through satellite DNA accumulation in *Megaleporinus elongatus*. *Cytogenet. Genome Res.* 160 (1), 38–46. doi:10.1159/000506265

da Silva, E. L., de Borba, R. S., and Parise-Maltempi, P. P. (2012). Chromosome mapping of repetitive sequences in Anostomidae species: implications for genomic and sex chromosome evolution. *Mol. Cytogenet.* 5 (1), 45. doi:10.1186/1755-8166-5-45

da Silva, M. J., Fogarin Destro, R., Gazoni, T., Narimatsu, H., Pereira dos Santos, P. S., Haddad, C. F. B., et al. (2020). Great abundance of satellite DNA in *Proceratophrys* (Anura, Odontophrynidae) revealed by genome sequencing. *Cytogenet. Genome Res.* 160 (3), 141–147. doi:10.1159/000506531

Dalíková, M., Zrzavá, M., Kubíčková, S., and Marec, F. (2017). W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res.* 25 (3), 241–252. doi:10.1007/s10577-017-9558-8

de Oliveira, R. R., Feldberg, E., Dos Anjos, M. B., and Zuanon, J. (2008). Occurrence of multiple sexual chromosomes (XX/XY_1Y_2 and $Z_1Z_1Z_2Z_2/Z_1Z_2W_1W_2$) in catfishes of the genus *Ancistrus* (Siluriformes: Loricariidae) from the Amazon basin. *Genetica* 134 (2), 243–249. doi:10.1007/s10709-007-9231-9

Deakin, J. E., Potter, S., O'Neill, R., Ruiz-Herrera, A., Cioffi, M. B., Eldridge, M. D. B., et al. (2019). Chromosomics: bridging the gap between genomes and chromosomes. *Genes* 10, 627. doi:10.3390/genes10080627

Devlin, R. H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208 (3), 191–364. doi:10.1016/S0044-8486(02)00057-1

dos Santos, R. Z., Calegari, R. M., Silva, D. M. Z. d. A., Ruiz-Ruano, F. J., Melo, S., Oliveira, C., et al. (2021). A long-term conserved satellite DNA that remains unexpanded in several genomes of Characiformes fish is actively transcribed. *Genome Biol. Evol.* 13 (2). doi:10.1093/gbe/evab002

Dover, G. A. (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet.* 2, 159–165. doi:10.1016/0168-9525(86)90211-8

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299 (5879), 111–117. doi:10.1038/299111a0

Drummond, A., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., et al. (2015). *Geneious v4. 8. 2009.* [Online]. Available: http://www.geneious.com (Accessed 10, 2020).

Dulz, T. A., Azambuja, M., Nascimento, V. D., Lorscheider, C. A., Noleto, R. B., Moreira-Filho, O., et al. (2020). Karyotypic diversification in two *Megaleporinus* species (Characiformes, Anostomidae) inferred from *in situ* localization of repetitive DNA sequences. *Zebrafish* 17 (5), 333–341. doi:10.1089/zeb.2020.1918

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5 (1), 113. doi:10.1186/1471-2105-5-113

Elder, J. F., and Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70 (3), 297–320. doi:10.1086/419073

Ellegren, H. (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat. Rev. Genet.* 12 (3), 157–166. doi:10.1038/nrg2948

Favarato, R. M., da Silva, M., de Oliveira, R. R., Artoni, R. F., Feldberg, E., and Matoso, D. A. (2016). Cytogenetic diversity and the evolutionary dynamics of rDNA genes and telomeric sequences in the *Ancistrus* genus (Loricariidae: Ancistrini). *Zebrafish* 13 (2), 103–111. doi:10.1089/zeb.2015.1140

Ferretti, A. B. S. M., Milani, D., Palacios-Gimenez, O. M., Ruiz-Ruano, F. J., and Cabral-de-Mello, D. C. (2020). High dynamism for neo-sex chromosomes: satellite DNAs reveal complex evolution in a grasshopper. *Heredity* 125 (3), 124–137. doi:10.1038/s41437-020-0327-7

Foresti, F., Toledo, A., and Toledo F°, S. A. (1981). Polymorphic nature of nucleolus organizer regions in fishes. *Cytogenet. Cel Genet* 31 (3), 137–144. doi:10.1159/000131639

Galetti Jr., P. M., Jr, Lima, N. R. W., and Venere, P. C. (1995). A monophyletic ZW sex chromosome system in *Leporinus* (Anostomidae, Characiformes). *Cytologia* 60 (4), 375–382. doi:10.1508/cytologia.60.375

Galetti, Jr., P. M., Jr, and Foresti, F. (1986). Evolution of the ZZ/ZW system in *Leporinus* (Pisces, Anostomidae). *Cytogenet. Cel Genet* 43 (1-2), 43–46. doi:10.1159/000132296

Garrido-Ramos, M. (2017). Satellite DNA: an evolving topic. *Genes* 8, 230. doi:10.3390/genes8090230

Godwin, J., and Roberts, R. (2018). "Environmental and Genetic Sex Determining Mechanisms in Fishes," in *Transitions Between Sexual Systems: Understanding the Mechanisms of, and Pathways Between, Dioecy, Hermaphroditism and Other Sexual Systems*. Editor J. L. Leonard (Cham: Springer International Publishing), 311–344. doi:10.1007/978-3-319-94139-4_11

Hashimoto, D. T., Parise-Maltempi, P. P., Laudicina, A., Bortolozzi, J., Senhorini, J. A., Foresti, F., et al. (2009). Repetitive DNA probe linked to sex chromosomes in hybrids between Neotropical fish *Leporinus macrocephalus* and *Leporinus elongatus* (Characiformes, Anostomidae). *Cytogenet. Genome Res.* 124 (2), 151–157. doi:10.1159/000207523

Kratochvíl, L., Stöck, M., Rovatsos, M., Bullejos, M., Herpin, A., Jeffries, D. L., et al. (2021). Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. *Phil. Trans. R. Soc. B* 376 (1833), 20200097. doi:10.1098/rstb.2020.0097

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Lorite, P., Carrillo, J. A., Tinaut, A., and Palomeque, T. (2004). Evolutionary dynamics of satellite DNA in species of the Genus *Formica* (Hymenoptera, Formicidae). *Gene* 332, 159–168. doi:10.1016/j.gene.2004.02.049

Lorite, P., Muñoz-López, M., Carrillo, J. A., Sanllorente, O., Vela, J., Mora, P., et al. (2017). Concerted evolution, a slow process for ant satellite DNA: study of the satellite DNA in the *Aphaenogaster* genus (Hymenoptera, Formicidae). *Org. Divers. Evol.* 17 (3), 595–606. doi:10.1007/s13127-017-0333-7

Louzada, S., Lopes, M., Ferreira, D., Adega, F., Escudeiro, A., Gama-Carvalho, M., et al. (2020). Decoding the Role of Satellite DNA in Genome Architecture and Plasticity-An Evolutionary and Clinical Affair. *Genes* 11, 72. doi:10.3390/genes11010072

Lower, S. S., McGurk, M. P., Clark, A. G., and Barbash, D. A. (2018). Satellite DNA evolution: old ideas, new approaches. *Curr. Opin. Genet. Dev.* 49, 70–78. doi:10.1016/j.gde.2018.03.003

Marreta, M. E., Faldoni, F. L. C., and Parise-Maltempi, P. P. (2012). Cytogenetic mapping of the W chromosome in the genus *Leporinus* (Teleostei, Anostomidae) using a highly repetitive DNA sequence. *J. Fish Biol.* 80 (3), 630–637. doi:10.1111/j.1095-8649.2011.03199.x

Molina, W. F., and Galetti, P. M. (2007). Early replication banding in *Leporinus* species (Osteichthyes, Characiformes) bearing differentiated sex chromosomes (ZW). *Genetica* 130 (2), 153–160. doi:10.1007/s10709-006-9002-z

Nakayama, I., Foresti, F., Tewari, R., Schartl, M., and Chourrout, D. (1994). Sex chromosome polymorphism and heterogametic males revealed by two cloned DNA probes in the ZW/ZZ fish *Leporinus elongatus*. *Chromosoma* 103 (1), 31–39. doi:10.1007/bf00364723

Nascimento, M., Sousa, A., Ramirez, M., Francisco, A. P., Carriço, J. A., and Vaz, C. (2017). PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* 33 (1), 128–129. doi:10.1093/bioinformatics/btw582

Negm, S., Greenberg, A., Larracuente, A. M., and Sproul, J. S. (2021). RepeatProfiler: a pipeline for visualization and comparative analysis of repetitive DNA profiles. *Mol. Ecol. Resour.* 21 (3), 969–981. doi:10.1111/1755-0998.13305

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29 (6), 792–793. doi:10.1093/bioinformatics/btt054

Ohta, T., and Dover, G. A. (1984). The cohesive population genetics of molecular drive. *Genetics* 108 (2), 501–521. doi:10.1093/genetics/108.2.501

Palacios-Gimenez, O. M., Dias, G. B., de Lima, L. G., Kuhn, G. C. e. S., Ramos, É., Martins, C., et al. (2017). High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*. *Sci. Rep.* 7 (1), 6422. doi:10.1038/s41598-017-06822-8

Palacios-Gimenez, O. M., Marti, D. A., and Cabral-de-Mello, D. C. (2015). Neo-sex chromosomes of *Ronderosia bergi*: insight into the evolution of sex chromosomes in grasshoppers. *Chromosoma* 124 (3), 353–365. doi:10.1007/s00412-015-0505-1

Palacios-Gimenez, O. M., Milani, D., Song, H., Marti, D. A., López-León, M. D., Ruiz-Ruano, F. J., et al. (2020). Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. *Genome Biol. Evol.* 12 (3), 88–102. doi:10.1093/gbe/evaa018

Parise-Maltempi, P., da Silva, E., Rens, W., Dearden, F., O'Brien, P. C., Trifonov, V., et al. (2013). Comparative analysis of sex chromosomes in *Leporinus* species (Teleostei, Characiformes) using chromosome painting. *BMC Genet.* 14 (1), 60. doi:10.1186/1471-2156-14-60

Parise-Maltempi, P. P., Martins, C., Oliveira, C., and Foresti, F. (2007). Identification of a new repetitive element in the sex chromosomes of *Leporinus elongatus*

(Teleostei: Characiformes: Anostomidae): new insights into the sex chromosomes of *Leporinus*. *Cytogenet. Genome Res.* 116 (3), 218–223. doi:10.1159/000098190

Pinkel, D., Straume, T., and Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci.* 83 (9), 2934–2938. doi:10.1073/pnas.83.9.2934

Plohl, M., Luchetti, A., Meštrović, N., and Mantovani, B. (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409 (1-2), 72–82. doi:10.1016/j.gene.2007.11.013

Plohl, M., Meštrovic, N., and Mravinac, B. (2012). Satellite DNA evolution. *Genome Dyn.* 7, 126–152. doi:10.1159/000337122

Poltronieri, J., Marquioni, V., Bertollo, L. A. C., Kejnovsky, E., Molina, W. F., Liehr, T., et al. (2014). Comparative Chromosomal Mapping of Microsatellites in *Leporinus* Species (Characiformes, Anostomidae): Unequal Accumulation on the W Chromosomes. *Cytogenet. Genome Res.* 142 (1), 40–45. doi:10.1159/000355908

Ramirez, J. L., Birindelli, J. L., Carvalho, D. C., Affonso, P. R. A. M., Venere, P. C., Ortega, H., et al. (2017a). Revealing hidden diversity of the underestimated neotropical ichthyofauna: DNA barcoding in the recently described genus *Megaleporinus* (Characiformes: Anostomidae). *Front. Genet.* 8, 149. doi:10.3389/fgene.2017.00149

Ramirez, J. L., Birindelli, J. L. O., and Galetti, P. M., Jr. (2017b). A new genus of Anostomidae (Ostariophysi: Characiformes): Diversity, phylogeny and biogeography based on cytogenetic, molecular and morphological data. *Mol. Phylogenet. Evol.* 107, 308–323. doi:10.1016/j.ympev.2016.11.012

Richard, G.-F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72 (4), 686–727. doi:10.1128/MMBR.00011-08

Rovatsos, M., Marchal, J. A., Giagia-Athanasopoulou, E., and Sánchez, A. (2021). Molecular composition of heterochromatin and its contribution to chromosome variation in the *Microtus thomasi/Microtus atticus* species complex. *Genes* 12, 807. doi:10.3390/genes12060807

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., and Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* 6, 28333. doi:10.1038/srep28333

Ruiz-Ruano, F. J., Navarro-Domínguez, B., Camacho, J. P. M., and Garrido-Ramos, M. A. (2019). Characterization of the satellitome in lower vascular plants: the case of the endangered fern *Vandenboschia speciosa*. *Ann. Bot.* 123 (4), 587–599. doi:10.1093/aob/mcy192

Schartl, M., Schmid, M., and Nanda, I. (2016). Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma* 125 (3), 553–571. doi:10.1007/s00412-015-0569-y

Schemberger, M. O., Bellafronte, E., Nogaroto, V., Almeida, M. C., Schühli, G. S., Artoni, R. F., et al. (2011). Differentiation of repetitive DNA sites and sex chromosome systems reveal closely related group in Parodontidae (Actinopterygii: Characiformes). *Genetica* 139 (11), 1499–1508. doi:10.1007/s10709-012-9649-6

Schemberger, M. O., Nascimento, V. D., Coan, R., Ramos, É., Nogaroto, V., Ziemniczak, K., et al. (2019). DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. *Chromosoma* 128 (4), 547–560. doi:10.1007/s00412-019-00721-9

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS ONE* 6 (3), e17288. doi:10.1371/journal.pone.0017288

Sember, A., Bertollo, L. A. C., Ráb, P., Yano, C. F., Hatanaka, T., de Oliveira, E. A., et al. (2018). Sex chromosome evolution and genomic divergence in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae). *Front. Genet.* 9, 71. doi:10.3389/fgene.2018.00071

Sember, A., Nguyen, P., Perez, M. F., Altmanová, M., Ráb, P., and Cioffi, M. d. B. (2021). Multiple sex chromosomes in teleost fishes from a cytogenetic perspective: state of the art and future challenges. *Phil. Trans. R. Soc. B* 376 (1833), 20200098. doi:10.1098/rstb.2020.0098

Serrano-Freitas, É. A., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Utsunomia, R., Araya-Jaime, C., Oliveira, C., et al. (2020). Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. *Mol. Genet. Genomics* 295 (1), 195–207. doi:10.1007/s00438-019-01615-2

Silva, D. M. Z. d. A., Utsunomia, R., Ruiz-Ruano, F. J., Daniel, S. N., Porto-Foresti, F., Hashimoto, D. T., et al. (2017). High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci. Rep.* 7 (1), 12726. doi:10.1038/s41598-017-12939-7

Smit, A. F. A., Hubley, R., and Green, P. (2013). *Repeat Masker. 2013*. [Online]. Available: http://www.repeatmasker.org/(Accessed 10, 2020).

Splendore de Borba, R., Lourenço da Silva, E., and Parise-Maltempi, P. P. (2013). Chromosome mapping of retrotransposable elements Rex1 and Rex3 in *Leporinus* Spix, 1829 species (Characiformes: Anostomidae) and its relationships among heterochromatic segments and W sex chromosome. *Mobile Genet. Elem.* 3 (6), e27460. doi:10.4161/mge.27460

Sproul, J. S., Khost, D. E., Eickbush, D. G., Negm, S., Wei, X., Wong, I., et al. (2020). Dynamic evolution of euchromatic satellites on the X chromosome in *Drosophila melanogaster* and the simulans clade. *Mol. Biol. Evol.* 37 (8), 2241–2256. doi:10.1093/molbev/msaa078

Terencio, M. L., Schneider, C. H., Gross, M. C., Nogaroto, V., de Almeida, M. C., Artoni, R. F., et al. (2012). Repetitive sequences associated with differentiation of W chromosome in *Semaprochilodus taeniurus*. *Genetica* 140 (10-12), 505–512. doi:10.1007/s10709-013-9699-4

Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, chromatin and evolution of satellite DNA. *Ijms* 22 (9), 4309. doi:10.3390/ijms22094309

Traut, W., and Winking, H. (2001). Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Res.* 9 (8), 659–672. doi:10.1023/A:1012956324417

Ugarković, Đ., and Plohl, M. (2002). Variation in satellite DNA profiles-causes and effects. *Embo J.* 21 (22), 5955–5959. doi:10.1093/emboj/cdf612

Utsunomia, R., Ruiz-Ruano, F. J., Silva, D. M. Z. A., Serrano, É. A., Rosa, I. F., Scudeler, P. E. S., et al. (2017). A glimpse into the satellite DNA library in Characidae fish (Teleostei, Characiformes). *Front. Genet.* 8, 103. doi:10.3389/fgene.2017.00103

Utsunomia, R., Silva, D. M. Z. d. A., Ruiz-Ruano, F. J., Goes, C. A. G., Melo, S., Ramos, L. P., et al. (2019). Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* 9 (1), 5856. doi:10.1038/s41598-019-42383-8

Volff, J.-N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity* 94 (3), 280–294. doi:10.1038/sj.hdy.6800635

Vondrak, T., Ávila Robledillo, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2020). Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J.* 101 (2), 484–500. doi:10.1111/tpj.14546

Wright, A. E., Dean, R., Zimmer, F., and Mank, J. E. (2016). How to make a sex chromosome. *Nat. Commun.* 7, 12087. doi:10.1038/ncomms12087

Yano, C. F., Bertollo, L. A. C., Ezaz, T., Trifonov, V., Sember, A., Liehr, T., et al. (2017). Highly conserved Z and molecularly diverged W chromosomes in the fish genus *Triportheus* (Characiformes, Triportheidae). *Heredity* 118, 276–283. doi:10.1038/hdy.2016.83

Yano, C. F., Bertollo, L. A. C., Liehr, T., Troy, W. P., and Cioffi, M. d. B. (2016). W Chromosome Dynamics in *Triportheus* Species (Characiformes, Triportheidae): An Ongoing Process Narrated by Repetitive Sequences. *Jhered* 107 (4), 342–348. doi:10.1093/jhered/esw021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

3.2 CAPÍTULO II: Comparative Analysis of Transposable Elements Dynamics in Fish with Different Sex Chromosome Systems

Carolina Crepaldi, Diogo Cavalcanti Cabral-de-Mello e Patricia Pasquali Parise-Maltempi

Artigo submetido para Genome (ISSN 1480-3321)

Manuscript ID: gen-2023-0134

Resumo em Português:

Elementos transponíveis (TEs) são componentes genômicos amplamente distribuídos com papéis significativos na evolução do genoma e diferenciação de cromossomos sexuais. Neste estudo, comparamos a composição de TEs de três peixes filogeneticamente relacionados com diferentes sistemas de cromossomos sexuais: *Megaleporinus elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$), *Megaleporinus macrocephalus* (ZZ/ZW) (ambos com cromossomos W altamente diferenciados) e *Leporinus friderici* (sem cromossomos sexuais heteromórficos). Criamos bibliotecas personalizadas de TEs para cada espécie usando métodos de *clustering* e anotação manual, e determinamos a dinâmica temporal dos TEs por meio de análise baseada em divergência. A abundância de TEs variou de 16 a 21% nos três mobilomas, sendo *L. friderici* o que apresentou menor quantidade geral. Os TEs prevalentes em todas as três espécies foram os transposons de DNA, que se dispersaram durante eventos recentes de proliferação. A atividade específica da espécie, no entanto, indicou eventos variáveis de expansão, especialmente para DNA/hATs e retrotransposons. Apesar de terem composições de TEs semelhantes, *Megaleporinus* e *Leporinus* têm histórias distintas de TEs que evoluíram após sua separação e o surgimento dos cromossomos sexuais, destacando sua rápida expansão de TEs ao longo de curtos períodos evolutivos.

Comparative Analysis of Transposable Elements Dynamics in Fish with Different Sex Chromosome Systems

Carolina Crepaldi[1,*], Diogo Cavalcanti Cabral-de-Mello[1] and Patricia Pasquali Parise-Maltempi[1,*]

Departamento de Biologia Geral e Aplicada, Instituto de Biociências, UNESP Universidade Estadual Paulista "Júlio de Mesquita Filho", Rio Claro, Brazil

*Correspondence: carolina.crepaldi@unesp.br, patricia.parise@unesp.br

Abstract

Transposable elements (TEs) are widespread genomic components with substantial roles in genome evolution and sex chromosome differentiation. In this study, we compared the TE composition of three closely related fish with different sex chromosome systems: *Megaleporinus elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$), *Megaleporinus macrocephalus* (ZZ/ZW) (both with highly differentiated W sex chromosomes), and *Leporinus friderici* (without heteromorphic sex chromosomes). We created custom TE libraries for each species using clustering methods and manual annotation and prediction, and we predicted TE temporal dynamics through divergence-based analysis. The TE abundance ranged from 16 to 21% in the three mobilomes, with *L. friderici* having the lowest overall. Despite the recent amplification of TEs in all three species, we observed differing expansion activities, particularly between the two genera. Both *Megaleporinus* recently experienced high retrotransposon activity, with a reduction in DNA TEs, which could have implications in sex chromosome composition. In contrast, *L. friderici* showed the opposite pattern. Therefore, despite having similar TE compositions, *Megaleporinus* and *Leporinus* exhibit distinct TE histories that likely evolved after their separation, highlighting a rapid TE expansion over short evolutionary periods.

Keywords: Transposon annotation; Genomic composition; Fish genomes; repetitive DNA; Transposable elements; Genome evolution

1. Introduction

Transposable elements (TEs) have long played a significant role in shaping eukaryotic genomes (Feschotte and Pritham, 2007; Belyayev, 2014; Sotero-Caio *et al.*, 2017; Bourque *et al.*, 2018). These elements, capable of inserting themselves into new genomic regions, contribute substantially to genetic diversity (Makałowski, 2000; Belyayev, 2014; Gao *et al.*, 2016; Sotero-Caio *et al.*, 2017; Yuan *et al.*, 2018). TEs can multiply and expand at variable rates due to their dynamic nature and ultimately influence host genome evolution (Feschotte and Pritham, 2007; Chalopin *et al.*, 2015a; Bourque *et al.*, 2018; Makałowski *et al.*, 2019; Carotti *et al.*, 2021).

TEs are widespread in fish, leading to a high degree of evolutionary plasticity in this speciose vertebrate group (Volff, 2005; Ahmad *et al.*, 2021; Carotti *et al.*, 2021; Etchegaray *et al.*, 2022). Even in compact fish genomes, TE proliferation accompanies rapid sequence turnover and significant structural heterogeneity (Schemberger *et al.*, 2019; Shao *et al.*, 2019; Carotti *et al.*, 2021). As a result, fish are great representatives of TE diversity and its impact on genomic evolution (Volff, 2005; Chalopin *et al.*, 2015a; Sotero-Caio *et al.*, 2017; Ahmad *et al.*, 2021). Advances in bioinformatics have facilitated the cataloging of TEs, enabling comparisons between mobilomes (the entire collection of mobile genetic elements inside a genome) in non-model species. Examining TE diversity and dynamics provide insights into various fish characteristics, including migratory behaviors, sex chromosome evolution, and even speciation (Chalopin *et al.*, 2015b; Schemberger *et al.*, 2019; Shao *et al.*, 2019; Ahmad *et al.*, 2021; Carotti *et al.*, 2021; Lehmann *et al.*, 2021; Etchegaray *et al.*, 2022).

Repetitive DNAs, including TEs and satellite DNAs (satDNAs), are known to play fundamental roles in sex chromosome evolution. These elements accumulate in non-recombining genome regions, which may cause chromosome rearrangements and influence sex chromosome heteromorphy (Charlesworth *et al.*, 1994; Chalopin *et al.*, 2015b; Schartl *et al.*, 2016; Wright *et al.*, 2016; Kent *et al.*, 2017; Charlesworth, 2019; Kratochvíl *et al.*, 2021). Unlike mammals and birds, fish exhibit diverse sex chromosome systems, such as XX/XY, ZZ/ZW, or multiple sex chromosomes from any heterogamety (Mank *et al.*, 2006; Schartl *et al.*, 2016; Wright *et al.*, 2016; Charlesworth, 2019). This diversity is even more striking when looking at closely related fish species, as even those within the same genus may have different sex chromosome systems (Desjardins and Fernald, 2009; Kitano and Peichel, 2012; Yano *et al.*, 2016a; De Freitas *et al.*, 2018; Sember *et al.*, 2018; Sember *et al.*, 2021). TEs are prevalent

in fish sex chromosomes, and have been found on both small and degenerated W chromosomes (Yano *et al.*, 2016b) as well as a giant Y sex chromosomes (Rosolen *et al.*, 2018). TEs can ultimately accelerate differentiation in relatively young sex chromosome systems (Kent *et al.*, 2017; Rosolen *et al.*, 2018; Dechaud *et al.*, 2019; Carducci *et al.*, 2020), presenting an additional layer of genomic complexity in fish (Charlesworth, 2019; Dechaud *et al.*, 2019; Schemberger *et al.*, 2019; Kratochvíl *et al.*, 2021).

The *Megaleporinus* genus, with highly differentiated and heterochromatic W sex chromosomes, includes *Megaleporinus elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$) and *Megaleporinus macrocephalus* (ZZ/ZW) (Galetti Jr and Foresti, 1986; Martins and Galetti, 1999; Venere *et al.*, 2004; Ramirez *et al.*, 2017a). In contrast, most species in the closely related genus *Leporinus* have undifferentiated sex chromosomes (homomorphic sex chromosomes). (Ramirez *et al.*, 2017a). This distinction provides an ideal scenario for investigating fish sex chromosome evolution and understanding variations in different sex systems (Parise-Maltempi *et al.*, 2013; De Barros *et al.*, 2018). *Megaleporinus* species heavily rely on the accumulation of repetitive DNA for sex chromosome differentiation (Parise-Maltempi *et al.*, 2007; Marreta *et al.*, 2012; Splendore De Borba *et al.*, 2013; Crepaldi and Parise-Maltempi, 2020). Ongoing genome research, with an emphasis on repetitive composition and evolution, could provide new insights into how the rapid formation of these distinctly differentiated sex chromosomes occurred (Utsunomia *et al.*, 2019; Crepaldi and Parise-Maltempi, 2020; Crepaldi *et al.*, 2021; Dos Santos *et al.*, 2021).

This study focuses on understanding fish repetitive DNA composition and differences in TE dynamics between species with different sex systems. We investigated TEs in three closely related fish species with distinct sex chromosome systems: *M. elongatus* ($Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$), *M. macrocephalus* (ZZ/ZW), and *Leporinus friderici* (non-identifiable sex chromosomes). We identified the TEs in these genomes employing short-read sequencing data and compared their evolutionary patterns by analyzing temporal activity and accumulation events. Our findings offer a detailed summary of TE composition and species-specific dynamics, contributing to a better understanding of the potential connection between TEs and sex chromosome differentiation in this group.

2. Materials and Methods

2.1 Animals and genome sequencing information

We selected two *Megaleporinus* species, with publicly available short-read sequenced reads, and the sole deposited *Leporinus* dataset for a comparative analysis with a close genus without heteromorphic sex chromosomes (Ramirez *et al.*, 2017a). The Illumina datasets from female *M. elongatus* (SRR25308063), female *M. macrocephalus* (SRR7263034) and *L. friderici* (SRR11676679) (unidentified sex) are accessible on NCBI GenBank.

2.2 Identification of TEs and construction of TE libraries

TEs were identified using RepeatExplorer2 (RE2) (Version 2.3.8.1) (available at https://repeatexplorer-elixir.cerit-sc.cz/galaxy/). We used several rounds of graph-based clustering analysis (Novák *et al.*, 2013) on the sequenced reads for each species to retrieve putative TEs. These prospected sequences were subsequently manually validated and curated to construct final TE consensus libraries for each analyzed species.

Firstly, we preprocessed the sequenced dataset from each species in the Galaxy platform (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/) using the "Quality trimming tool", "FASTQ interlacer on paired-end reads", "FASTQ to FASTA" tools. The generated FASTA file was then used as input in "RepeatExplorer2 clustering" with the standard recommended parameters. We followed preprocessing with several clustering rounds for each sequenced dataset, filtering out reads matching previously assembled outputs from RepeatExplorer2 with *deconseq* 0.4.3 (Schmieder and Edwards, 2011). Filtering out these previously identified putative repetitive elements helped avoid redundancy and ensured that subsequent iterations retrieved sequences with lower abundance. Highly abundant contigs tend to reappear in successive clustering rounds if not eliminated. Nevertheless, we analyzed and annotated all subsequent contigs before restarting the process of filtering and clustering.

At the end of the clustering process, we had several RE2 output files with repetitive clusters for each species. These were then manually examined for TE candidates and followed by consensus curation. We analyzed the RE2 outputs, paying closer attention to putative repetitive elements with a linear graph structure in the resulting clusters (Novák *et al.*, 2013; Novák *et al.*, 2020) and to the built-in RepBase annotation results (RE2 performs searches against RepBase during clustering, which sometimes can return positive matches to known TEs).

The dense ring-shaped graph shapes from RE2 output are usually deemed putative satDNAs; however, they also were manually examined due to the occurrence of putative LRT

elements also presenting dense graph topologies. We estimated the repeat abundances for the remaining repeats, mainly satDNAs, as RE2 also gives percentage values for all identified repeat sequences in the sequenced reads. For *Megaleporinus,* we also compared these values to previously reported satDNA data (Utsunomia *et al.*, 2019; Crepaldi and Parise-Maltempi, 2020; Crepaldi *et al.*, 2021). All repetitive sequences deemed not-TEs were excluded from the final curated libraries, serving mainly to avoid under or overestimation of TEs in case some output results from RE2 mistakenly included TEs as other repetitive DNA.

Overall, results from RE2 were presented as separate contigs for each putative repetitive element (designated as clusters). These internal contig structures were aligned using MUSCLE (Edgar, 2004), through Geneious v.4.8.5 (Drummond *et al.*, 2015), and used to construct the consensus sequences for each putative TE to subsequent confirmation and categorization. into known DNA transposons or retrotransposons through homology-based comparison with known TE databases.

For this step, each consensus was compared against previously reported TE profiles using the GiriCensor tool (Kohany *et al.*, 2006) from Repbase (Update 20221227) (Jurka *et al.*, 2005) and the Dfam database (version 3.6) (Storer *et al.*, 2021). We took a conservative approach to annotation and sequence comparison, looking for hallmarks of specific TE subclasses such as LTRs, and implementing an 80% similarity threshold in Repbase and Dfam results, adhering to the 80-80-80 rule (Wicker *et al.*, 2007; Flutre *et al.*, 2011). Sequences were considered a positive match and belonging to the previously known TEs if more than 80% of the query coverage could be aligned to resulting subjects (database results) with over 80% of identity (Wicker *et al.*, 2007). Sequences with short fragments less than 80% similar to known TEs were considered incomplete representatives and initially designated as unclassified transposons.

TEs within the same superfamily might exhibit differences, especially interspecies, and considering our homology-based approach compared the generated consensus to sequences from online databases. Additionally, our conservative approach to homology-based searches may result in the loss of putative TEs that were filtered out in the previous step. As such, to bypass these constraints, we also translated and searched for TE protein domains (integrase, reserve transcriptase and transposase) using Pfam (Mistry *et al.*, 2021), Swiss-Prot/UniProt (Wang *et al.*, 2021) and HMMER (version (Finn *et al.*, 2011; Potter *et al.*, 2018). This also aided in classifying some of the aforementioned unclassified and unknown TEs. Sequences

with an e-value lower than 1e-5 were reintegrated in the TE libraries. Final prediction of TE cataloguing into classes, subclasses, superfamilies and families followed classification by Bourque *et al.* (2018).

In the final step, all curated consensus sequences were checked for redundancies using CD-HIT (Fu *et al.*, 2012) also based on the 80-80-80 rule (Wicker *et al.*, 2007; Flutre *et al.*, 2011). This was performed individually for each library. Final custom libraries for *M. elongatus*, *M. macrocephalus* and *L. friderici* were individually merged with publicly available TE libraries from *Astyanax mexicanus* and *Danio rerio* from FishTEDB.

2.3 TE quantification, divergence, and insertion time estimation

The three manually curated TE libraries, integrated with *A. mexicanus* and *D. rerio* TEs, were used as inputs for TE quantification and sequence divergence analysis. These analyses were carried out independently for each genome using RepeatMasker (Smit *et al.*, 2017) and a custom script available at https://github.com/fjruizruano/satminer/blob/master/repeat_8masker_run_big.py.

The RepeatMasker tool calcDivergenceFromAlign.pl estimated Kimura-2-Parameter (K2P) values (nucleotide divergence values) for each TE and their corresponding genomic abundance. The predicted consensus sequences from each library provided ancestral TE estimations, and K2P determines the divergence between individual TE copies in the genomes and the corresponding consensus sequence (Smit *et al.*, 2017). Using the RepeatMasker divergence summary data, we calculated the abundance by dividing the nucleotide proportions for each TE by the library size (in total base pairs). Then, we created RepeatLandscapes for each species using R package version 4.1.2 (R Core Team, 2021).

Finally, we established TE insertion time and activity dates using the previously indicated K2P divergence summary data from RepeatMasker. K2P values were converted to MYA using the formula T = k/2r (Li *et al.*, 2019), where “T” is the estimated age (MYA), “k” is the pairwise divergence values, and “r” is the nucleotide substitution rate. The value “r” was determined as the neutral mutation rate applied to transposons overall, assumed to be an average of 10^{-8} substitutions per synonymous site per year for TEs (Arkhipova and Meselson, 2005; Schemberger *et al.*, 2019). We applied the same formula for all three genomes, given the substitution rate was a general value for all transposons analyzed in the present work and lack of mutation rate values for the species.

3. Results

3.1 Repetitive DNA composition and TE diversity

We analyzed the transposable element (TE) libraries of three fish species: female *M. elongatus*, female *M. macrocephalus*, and a non-sexed *L. friderici*. Figure 1 and Online Resource 1 illustrate that all three species have rather similar repetitive DNA amounts, including TEs, satDNAs, and other unidentified repeats constituting about one-third of their genomes (Figure 1a).

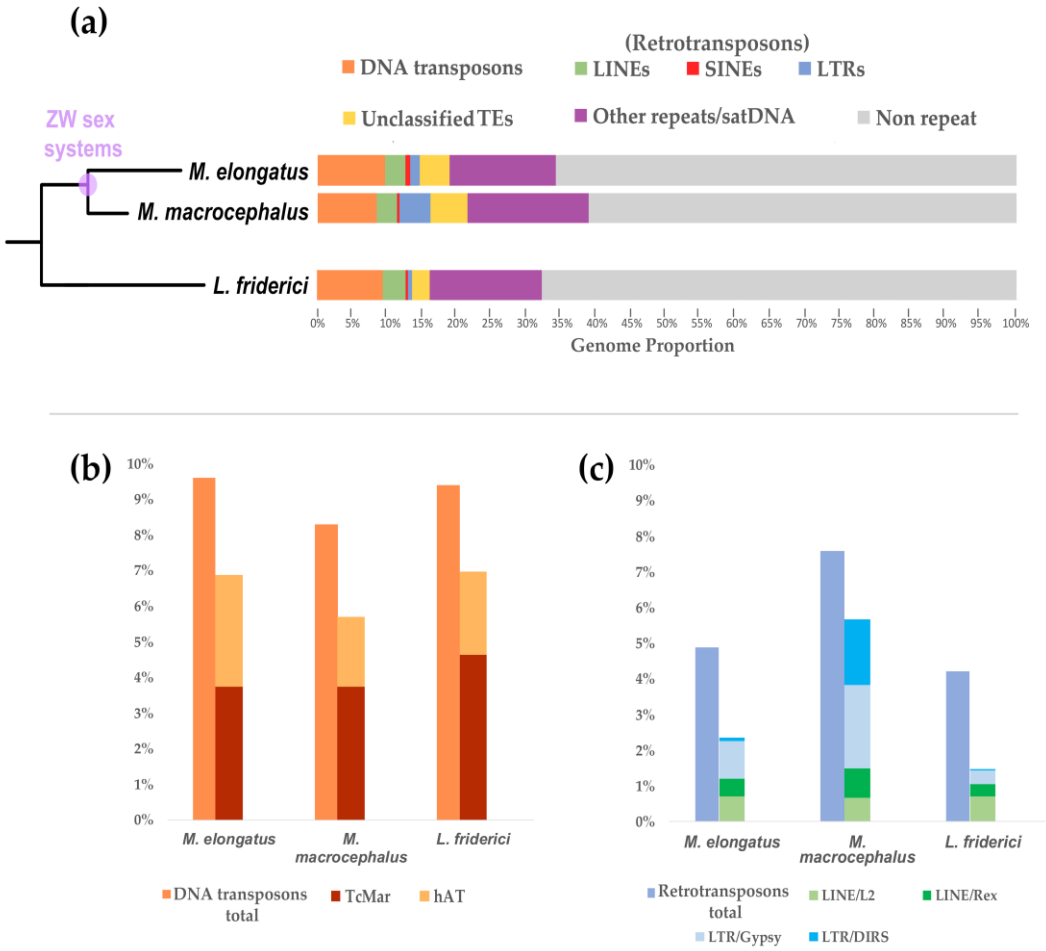


Fig. 1 (a) Estimated proportions of transposable elements (TEs) (DNA transposons, retrotransposons, unclassified TEs) and other repeats for the analyzed sequenced reads of female *Megaleporinus elongatus*, female *Megaleporinus macrocephalus*, and *Leporinus friderici*, displayed on a simplified phylogenetic tree based on Ramirez *et al.* (2017a). The purple highlight indicates the emergence of heteromorphic sex chromosomes (ZW systems), which are absent in *Leporinus friderici*. (b) Percentage of the most abundant DNA transposon superfamilies, hAT and TcMariner, in each species compared to the total DNA transposon content. (c) Percentage of the most abundant retrotransposon families in relation to overall retrotransposons.

As we focused on transposable elements, the analyzed reads for each species yielded a total of 61 different TE elements, including 33 DNA transposons and 28 retrotransposon (Online Resource 1). While the majority of TEs were shared among the species, there were exceptions, such as specific absent elements in *M. macrocephalus* (LINE Mir, Proto2, R2-hero) and hAT families hAT-hAT5 and hAT-hAT6, with these last two also absent in *L. friderici* (Online Resource 1). Despite being detected in *M. elongatus*, these two hAT families also had very low abundance in this species.

TE content was approximately 16% in *L. friderici* and roughly 19% and 21% in *M. elongatus* and *M. macrocephalus*, respectively (Figure 1a). Despite the diversity of TE elements, few dominant ones (Figures 1b, 1c) played a significant role in abundance, contributing to genetic differences among species. DNA transposons was the most common Class of TEs, accounting for about 8% (*M. macrocephalus*), to around 9% (*L. friderici*), and around 10% (*M. elongatus*) (Figure 1a, Online Resource 1). TcMariner and hAT were the two dominant superfamilies in all three species (Figure 1b, Online Resource 1). These two superfamilies alone accounted for approximately 6% to 7% of total genome composition, a sizeable fraction given that the DNA transposon class account for nearly 10% of the genomes as whole.

TcMariner, encompassing four families (Tc1, TcMar-ISRm11, TcMar-Tc2, and TcMar-Tigger), was the most common TE superfamily across species (Figure 1b, Online Resource 1), accounting for 4.63% of the *L. friderici* genome and 3.74% of both *Megaleporinus* genomes. The equally widespread hAT superfamily was also prevalent (families Ac, Blackjack, Charlie, Tip100, and several elements within these families, hAT-hAT5, hAT-hAT6, hAT-hATm and hAT-Tol2) (Online Resource 1). The most prevalent hAT elements are not annotated (not assigned to a specific family). *M. elongatus* had the highest incidence of hAT elements (3.12%) among the species, followed by *L. friderici* (2.34%) and *M. macrocephalus* (1.95%).

Retrotransposons (LINEs, SINEs, and LTRs) were less abundant but exhibited increased divergence bias towards *Megaleporinus* (species with differentiated sex chromosomes) (Figures 1a, c). Retrotransposons accounted for 4.3% of *L. friderici's* total genome size (3% of LINEs, 1% of LTRs, and 0.3% of SINEs), 5% of *M. elongatus'* genome (3% of LINEs, 1% of LTRs, and 1% of SINEs), and 7.3% of *M. macrocephalus'* genome (3% of LINEs, 4% of LTRs, and 0.3% of SINEs) genomes. The high prevalence of LTRs Gypsy and DIRS in *M. macrocephalus* (Figure 1c) most likely influenced these differences. The most prevalent LINE retroelements, namely L2 and Rex, were also more abundant in *M.*

macrocephalus. In contrast, they accumulated in a roughly similar pattern in *M. elongatus* and *L. friderici* (Figure 1b, Online Resource 1). SINEs were rare in all three genomes, with the SINE derived from transfer RNA (tRNA) being the most prevalent, accounting for 0.232% of the *M. macrocephalus* genome, 0.298% of *L. friderici,* and 0.530% of *M. elongatus* (Online Resource 1).

Despite successful identification of many TEs, some remained unannotated, comprising 2.43% of *L. friderici*, 4.26% of *M. elongatus*, and 5.17% of *M. macrocephalus* genomes. These unannotated TEs were either undetected by RepeatExplorer2 or RepeatMasker or did not match positively in Repbase, Dfam, or protein domain searches using the 80-80-80 rule. Other repetitive elements, including satDNAs and unclassified repetitive sequences, accounted for 15% in *M. elongatus*, 16% in *L. friderici,* and 17% in *M. macrocephalus* (Figure 1a, Online Resource 1) and added to their total amount of repetitive DNA content.

3.2 Interspecies TE dynamics and expansion events

After characterizing the TE composition of the three analyzed species, our subsequent focus was on understanding how these sequences expanded within these genomes, and whether their dynamics also exhibit similarities, akin to their diversity. We determined the putative insertion time (MYA) and transposition events for the two main classes of TEs, DNA transposons and retrotransposons (Figure 2), in *M. elongatus*, *M. macrocephalus* and *L. friderici*. By converting K2P divergence values to MYA, we applied neutral mutation rates for transposon sequences overall (Li *et al.*, 2019; Schemberger *et al.*, 2019). Transposition events preceding 25 MYA could not be recovered, likely due to copy loss, degeneration or limitations in our time estimation analysis.

The violin plot (Figure 2) suggests that TE expansion overall was predominantly recent, starting mainly around 15 MYA. Both *Megaleporinus* species and *L. friderici* exhibit a trend of recent transposition for both DNA transposons (Figure 2a) and retrotransposons (Figure 2b). *M. macrocephalus,* shows a significant increase in retrotransposons compared to the other species (indicated darker purple dot a the base of the plot on Figure 2b), possibly due to the prevalence of DIRS and LTR Gypsy in its genome (Figure 1c). Overall, all three species have undergone recent transposition events involving mainly young TE families.

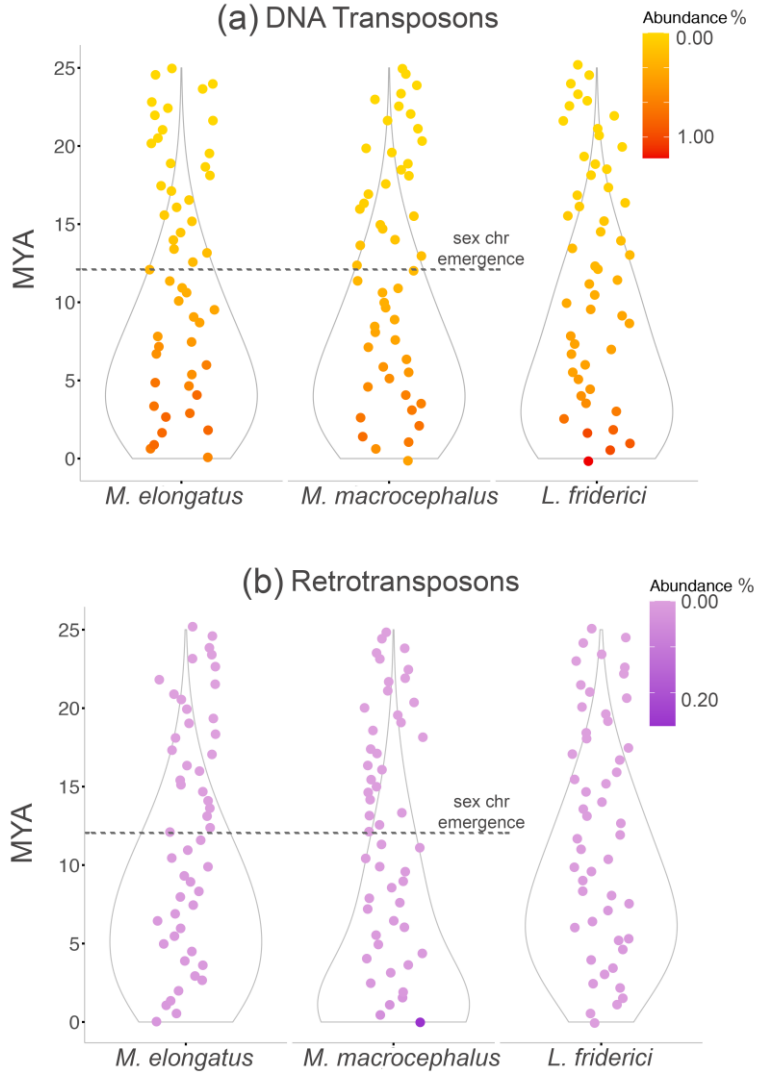


Fig. 2 Violin plots illustrating the estimated insertion time (MYA) for the main TE classes, (a) DNA transposons and (b) retrotransposons, in the examined species. Colored dots represent different distributions of concatenated abundance values of all elements for each Class of TE, along an ascending gradient (darker colored points indicate higher abundance values for that TE type). For reference, the dotted line depicts the predicted emergence of the heteromorphic sex chromosomes in *Megaleporinus elongatus* and *Megaleporinus macrocephalus*, given the phylogeny from Ramirez *et al.* (2017b).

We also evaluated this dynamism by analyzing Kimura-2-Parameter (K2P) divergence landscape graphs for both TE classes (Figure 3). The landscape topology shows an increase in TE copies within the genomes primarily through new insertions, reflecting recent transposition similar to Figure 2. Low divergence values across the landscapes (found on the leftmost side of the graph) indicate an recent expansion phase for all three species (Figure 3, Online Resources

2-4), with most copies exhibiting a K2P-value of <10, some at K-values 10–20, and gradually decreasing in abundance as divergence exceeds k = 20 (Figure 3).

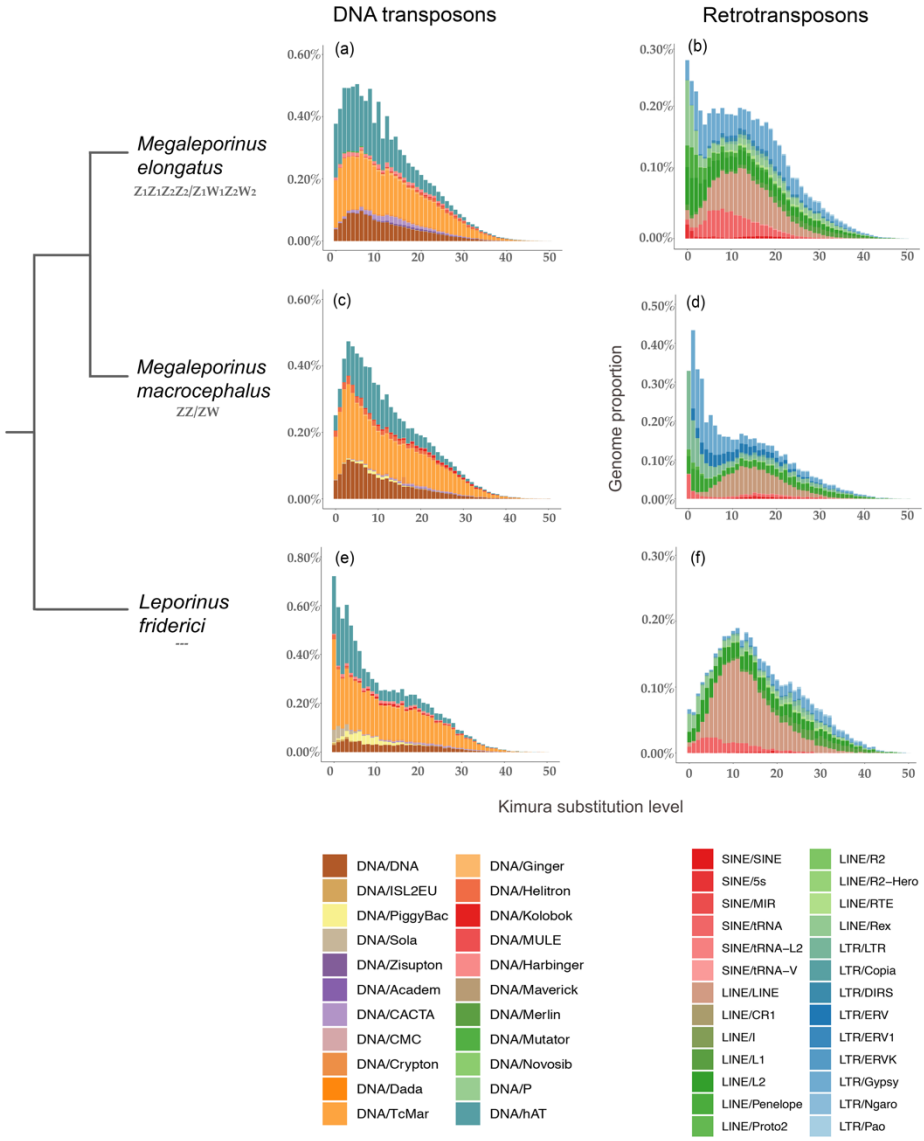


Fig. 3 K2P divergence analysis of species-specific TE elements from *Megaleporinus elongatus* (a, b), *Megaleporinus macrocephalus* (c, d), and *Leporinus friderici* (e, f). The phylogenetic relationships on the left are based on Ramirez *et al.* (2017a). The graphs on the right depict the landscape plots obtained by divergence analysis (x-axis) in relation to the genome proportion for each element type (y-axis) for the three studied species, separated by TE class, DNA transposons (a, c, e) and retrotransposons (b, d, f). It is worth noting that not all y-axis scales (genome proportion) are the same, especially for retrotransposons, which are smaller. The copies depicted on the left side of the graphs show minimal divergence from the consensus

sequences, indicating recent insertions. In contrast, the right side of the landscape plots depicts older clustering events at higher K2P values.

Despite overall recent activity as a whole, there are contrasting expansion patterns between the genera for the two main TE classes.

Firstly, DNA transposon invasions mainly involved hATs, TcMariners, and unclassified DNA transposons (referred to as DNA/DNA) for all three species (Figures 3a, c, e). Notable distinctions include a continuous and abundant hAT transposition event in *M. elongatus*, spanning from k-values 0 to 15, but with ongoing decline in activity, which was also experienced by *M. macrocephalus.* Conversely, *L. friderici* exhibits a recent and sudden increase in active DNA TEs.

As for retrotransposons, an opposite trend was observed. Both *Megaleporinus* had two major retrotransposon invasion phases (Figures 3b, d, f), with recent activity of L2, Rex, Gypsy, and tRNA elements. Additional expansion of 5s and Penelope can also be seen in *M. elongatus*, and prominent LTRs were recently active in *M. macrocephalus*. In contrast, *L. friderici* showed overall lower retrotransposon abundance and no notable peaks in recent events (<5 MYA) (Figure 3f).

Considering these contrasting dynamics, we highlighted the most widespread TE elements in all three species, with individual line plots for each TE (Figure 4). It is worth noting that no superfamily appeared to have reached a plateau during accumulation, indicating ongoing fluctuations in activity for both genera (Figure 4).

For DNA transposons (Figure 4a, 4b), TcMariners (Figure 4a) initially displayed a consistent and gradual accumulation across all three species, followed by a recent and sudden plummet for *Megaleporinus* (<5MYA). This contrasts with the steady increase in *L. friderici*, peaking in very recent activity (<5 MYA) (Figure 4a). Regarding hATs (Figures 4b), both *Megaleporinus* species show an unusual pattern of insertions between 5 and 10 MYA. *M. elongatus* displays numerous high peaks starting at 10 MYA, with several accumulation events followed by sudden declines. *M. macrocephalus* shows a similar pattern, but with a less expressive increase in abundance at each event. Despite this initial expansion history, both *Megaleporinus* species experienced an ongoing decline in recent hAT activity, mirroring the pattern observed for TcMariner. In contrast, once again, *L. friderici* had an expressive and very recent expansion for hATs, starting at 5 MYA, after low accumulation at older events.

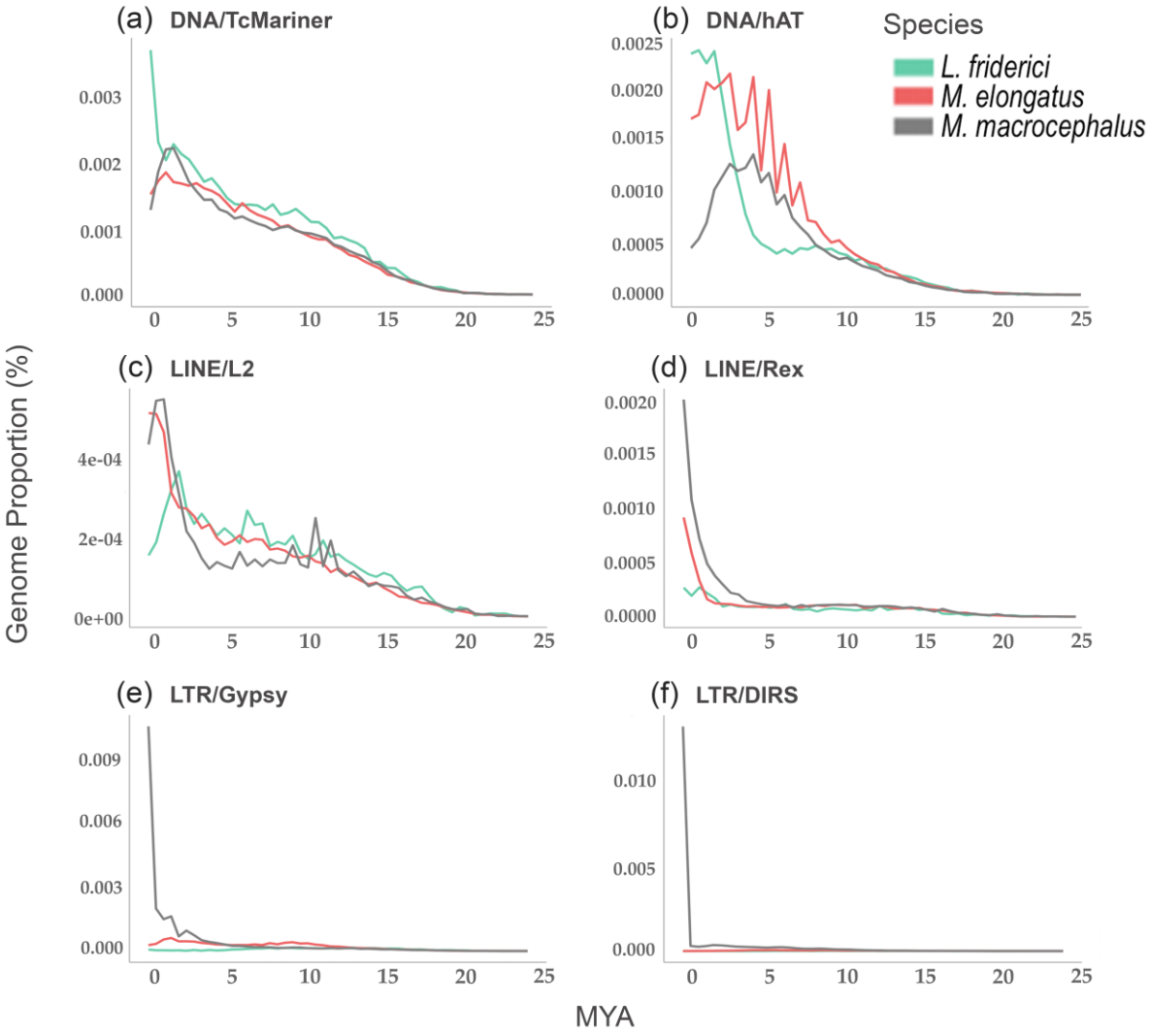


Fig. 4 Line color-coded graphs depicting the activity and temporal accumulation of the most enriched TE elements in female *Megaleporinus elongatus*, female *Megaleporinus macrocephalus*, and *Leporinus friderici* (a,b) DNA transposons and (c-f) retrotransposons. The x-axis represents time in millions of years (MYA) as established through divergence analysis, and the y-axis indicates the corresponding genomic proportion of TE copies for each species. Note the differences in abundance (%) for each element. The copies clustering on the leftmost side indicate possible recent insertion events.

As for retrotransposons, the earlier invasion phases primarily involved L2 copies (Figure 4c). Overall, transposition rates were similar among species until <5 MYA, at which point a considerably higher activity was observed in *M. macrocephalus* and *M. elongatus*; compared to the decline in *L. friderici*, with modest and scattered proliferation for most of the predominant retrotransposons.

The other prevalent retrotransposons (Rex, Gypsy, and DIRS) (Figure 4d, 4e, 4f) exhibited the most significant activity in *M. macrocephalus* in very recent timeframes. *M.*

elongatus also experienced ongoing Rex expansion (1–5 MYA), although with fewer cumulative events than its congeneric counterpart (Figure 4d). Gypsy and DIRS in particular showed substantial accumulation only in *M. macrocephalus* (Figure 4e, 4f).

4. Discussion

Mobile elements play a crucial role in shaping fish genomes (Carducci *et al.*, 2020), although the extent of their evolution and impact in the group remains unclear (Shao *et al.*, 2019; Carotti *et al.*, 2021). Here, we investigated transposons in fish with diverse sex chromosome systems, focusing on TE composition, expansion features and possible differences among species.

Megaleporinus and *Leporinus*, though geographically distant genera, separated recently (in the last few million years), and *M. elongatus* and *M. macrocephalus* shared a last common ancestor around 12 MYA (Ramirez *et al.*, 2017a). The evolutionary link between these two genera is reflected in the overall similarities of their TE diversity uncovered in the present work, another indication that closely related species are more likely to have a comparable mobilome composition (Shao *et al.*, 2019; Carotti *et al.*, 2021).

Unannotated TEs were also found in all analyzed genomes, which may represent incomplete sequences or lineage-specific components not detected by existing methods. Their fragmented nature likely results from not being recovered as complete units, and the failure to annotate these sequences may also be attributed to the conservative approach commonly used in manual curation (Bell *et al.*, 2022) and the limitations of *de novo* methods processes as of date.

While there were no explicit differences in repetitive content across species, both *Megaleporinus* showed slightly higher TE abundance than *L. friderici* (19-21% for the former against 16% for the latter). This could be attributed to the high heterochromatic content of their sex chromosomes, known to abundantly accumulate repetitive DNA (Parise-Maltempi *et al.*, 2013; Chalopin *et al.*, 2015b; Wright *et al.*, 2016).

TcMariners and hATs were the predominant TEs overall, followed by the LINEs L2 and Rex, and these four superfamilies are actively mobilized across multiple fish genomes (Valente *et al.*, 2010; Ellegren, 2011; Carducci *et al.*, 2018; Chen *et al.*, 2018; Yuan *et al.*, 2018; Shao *et al.*, 2019). Their rapid expansion and widespread distribution over other young, low-copy TEs (Figures 1b, c), possibly indicate their evolutionary success in proliferation and

potential impact on genome structure (Belyayev, 2014; Coan and Martins, 2018; Shao *et al.*, 2019; Srikulnath *et al.*, 2022).

All three species experienced recent transposon expansion waves, a regular occurrence in fish genomes (Chalopin *et al.*, 2015b; Schemberger *et al.*, 2019; Shao *et al.*, 2019). However, their patterns of copy accumulation of these shared TE superfamilies have species-specific dynamics as they have expanded differently among *Megaleporinus* and *Leporinus*, a feature that could alter specific genomic patterns as species diverge (Renaut and Bernatchez, 2011; Rogers *et al.*, 2018; Serrato-Capuchina and Matute, 2018; Mérot *et al.*, 2020).

Despite the similar occurrence of young families in all three species, a contrasting pattern of TE invasions can be seen between the analyzed genera. Both *Megaleporinus* show a reduction in DNA TE activity alongside very recent active retrotransposons; contrasting with *L. friderici*, with exhibits few active retrotransposons and higher DNA TE activity. Considering this discrepancy between genera with different sex chromosome systems, these dynamics in active TE in *Megaleporinus* could have implications in their highly heterochromatic sex chromosomes and warrant further investigation.

The dynamics of Rex elements, for example, have shown increased expansion in both *M. elongatus* and *M. macrocephalus* in recent years, differing from the much tamer accumulation in *L. friderici*. Interestingly, this family is known to be involved in sex chromosome differentiation (Yano *et al.*, 2016b; Carducci *et al.*, 2019; Carducci *et al.*, 2020), and analyzed Rex sequences formed distinct clusters in the W chromosomes of *Megaleporinus* (Splendore De Borba *et al.*, 2013; Dulz *et al.*, 2019).

As for hATs, there are similarities of small fragments with a satDNA found in *M. elongatus* (MelSat01-36) (Crepaldi and Parise-Maltempi, 2020) and its identical counterpart (MmaSat06-36) in *M. macrocephalus* (Utsunomia *et al.*, 2019). These satDNAs were highly clustered in the heteromorphic sex chromosomes, albeit in different patterns between *Megaleporinus*, while there was no positive physical mapping in *L. friderici*. Because satDNAs can derive from mobile elements, sequence similarities frequently indicate an interaction between these repetitive groups (Meštrović *et al.*, 2015; Carducci *et al.*, 2019; Paço *et al.*, 2019). Further exploration of this interaction might yield additional details about their expansion within the group, and more specifically on the sex chromosomes.

The sex chromosomes (W in *M. macrocephalus* and W_1 *M. elongatus*) share a common origin (Marreta *et al.*, 2012) and likely inherited transposon sequences from an ancestral species. Both species showed variations in TE dynamics as well, particularly the unusual hAT expansion in *M. elongatus* and the highly active retrotransposons (mainly DIRS and Gypsy) in

M. macrocephalus. Along with several other repetitive elements that have species-specific expansion (Splendore De Borba *et al.*, 2013; Poltronieri *et al.*, 2014; Dulz *et al.*, 2019; Utsunomia *et al.*, 2019; Crepaldi and Parise-Maltempi, 2020; Crepaldi *et al.*, 2021), these differing TE patterns reinforced this overall sequence dynamism in the group.

In summary, our study highlights that variations in transposition histories resulted in different mobilome dynamics between two genera with startlingly different sex chromosomes *Megaleporinus* and its sister genus, *Leporinus*, despite their similar composition and TE diversity. These changes likely occurred over relatively short evolutionary timeframes and might have contributed to the structural genomic variations in their distinct sex chromosome systems. More investigation is necessary to establish the time correlation between TE insertions and sex chromosome emergence, focusing on using specific mutation rates for each species, given the limitations of our current estimates. Chromosome-level assemblies and identification of TE insertion sites could also further highlight their correlation with sex chromosome composition.

Future research could also explore beyond short-read data, utilizing genome assemblies when available, as well as TE libraries at even higher resolution, with focus on specific TE families and their discrepancies between species. Additionally, physical mapping and co-localization assays with satDNAs, as noted before, could shed light on further composition details in sex chromosomes. These approaches, along with TE libraries of other *Megaleporinus* and *Leporinus* species, could provide deeper insights into the mechanisms driving TE evolution and their accumulation history within the group.

References: *The references used in this chapter are presented at the end of this thesis.*

Supplementary Material: *Available at: https://drive.google.com/drive/folders/1bfawekjikxrRUOd7XDVKa1SMT-Og3C6e?usp=share_link*

Author Contributions: C.C. conceptualized the work, alongside methodology definition, data acquisition, analysis and interpretation, and wrote the manuscript. D. C. C. M. contributed to data interpretation, manuscript writing, and revision. P.P.P.-M. funded and coordinated the research, as well as contributing to the writing and revision of the manuscript.

Acknowledgments: The authors would like to thank the members of the Animal Cytogenetics Laboratory (UNESP) for their valuable suggestions and constructive comments, especially MSc. Marcelo J. da Silva. We are grateful to Dr. Diogo Milani for his valuable suggestions and help with the bioinformatic analyses. We also thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) for funding this research.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #2017/00195_7, FAPESP #2018/12906-8, and FAPESP #2019/19069-7).

Ethics Approval. The animal study protocol was approved by the Animal Ethics Committee (CEA) of Sao Paulo State University (protocol number 3524, approval code 09/2017).

Data Availability: The genome datasets presented in this study are in publicly available repositories (NCBI) under accession numbers SRR25308063 (female *M. elongatus*), SRR7263034 (female *M. macrocephalus*) and SRR11676679 (*L. friderici).*

Conflicts of Interest: The authors declare no conflict of interest.

4. CONSIDERAÇÕES FINAIS

A presente tese reúne dados relacionados à composição repetitiva genômica entre espécies de peixes próximas com sistemas de cromossomos sexuais distintos no intuito de investigar a influência desta porção repetitiva e sua dinâmica na formação dos cromossomos sexuais entre estas espécies. Os dados geraram dois manuscritos que fornecem informações sobre a composição repetitiva e a evolução genômica em espécies destes peixes neotropicais. Os resultados aqui apresentados evidenciam a riqueza e a diversidade de DNAs repetitivos em *Megaleporinus* e *Leporinus*, destacando seu acúmulo recente nos genomas das três espécies estudadas. Apesar de *Megaleporinus* compartilharem satDNAs e *Megaleporinus* e *Leporinus* apresentarem mobilomas com diversidade de famílias semelhantes, os conteúdos repetitivos exibem dinâmicas específicas para cada espécie, especialmente em relação às famílias mais abundantes. A variabilidade na abundância, com acúmulo diferenciado em cromossomos sexuais (no caso de satDNAs – CAPÍTULO I) e eventos de inserção específicos (para TEs – CAPÍTULO II), evidenciam a organização altamente plástica destas sequências, desenvolvendo caminhos evolutivos independentes apesar da proximidade filogenética entre as espécies.

No contexto específico do cromossomo sexual heteromórfico de *M. elongatus* (CAPÍTULO I), a prevalência de apenas algumas famílias específicas e altamente abundantes na fêmea se mostrou um dos principais impulsionadores do acúmulo no cromossomo sexual heterocromático. Contrariando nossas expectativas, a fêmea não apresentou a maior diversidade de famílias de satDNAs entre os sexos; pelo contrário, o macho também exibiu famílias de satDNA enviesadas para seu sexo. Porém, com cópias muito raras e pouco acumuladas no genoma, os satDNA do macho não se mostraram altamente acumulados como visto na fêmea. Com isso, os principais contribuintes para a alta abundância de satDNAs no cromossomo sexual W_1 e possivelmente sua diferenciação foram a amplificação e a homogeneização das famílias mais abundantes da fêmea, promovendo a formação de grandes arranjos ao longo do cromossomo sexual heteromórfico.

Quanto ao conteúdo de elementos transponíveis (CAPÍTULO II), constatamos que a ocorrência e diversidade de TEs entre as espécies de *Megaleporinus*, com cromossomos sexuais heteromórficos, e *L. friderici* são relativamente semelhantes. São apenas alguns casos específicos de maior incidência de famílias em *Megaleporinus*, no geral, as famílias de TEs mais abundantes em seus genomas. As famílias mais prevalentes, como os transposons de DNA hAT e Tc-Mariners apresentaram dinâmicas espécie-específicas de inserção e acúmulo nas

espécies, com diferenças inclusive entre as duas espécies de *Megaleporinus*, e retrotransposons do tipo LTR se mostraram em abundância especificamente em *M. macrocephalus*. Assim, semelhante ao que constatamos para satDNAs, as maiores diferenças estão na dinâmica de acúmulo de determinadas famílias, e não necessariamente na diversidade de TEs encontrada para cada espécie.

As informações aqui geradas, em conjunto com estudos de genômica de outros vertebrados na literatura e aqueles desenvolvidos em tandem à esta tese, reiteram peculiaridades sobre elementos repetitivos em peixes, especialmente sua variabilidade e impacto na dinâmica de um genoma, além do suporte para a diferenciação de cromossomos sexuais heteromórficos.

Esforços futuros podem ser direcionados em incluir mais espécies de *Megaleporinus* e *Leporinus* expandindo o uso de dados genômicos, a fim de (a) traçar um histórico evolutivo destas sequências para melhor compreender mecanismos evolutivos como a hipótese de biblioteca de satDNAs; (b) integrar informações moleculares com dados de localização cromossômica também para transposons, em especial com foco nos mais proeminentes em fêmeas com cromossomos sexuais heteromórficos; (c) buscar marcadores genéticos de determinação do sexo por meio de outras técnicas, como o uso de sequenciamento RadSeq (*Restriction site associated DNA sequencing*), já atualmente utilizado para busca de marcadores sexuais em espécies não-modelo especialmente de peixes e anfíbios.

5. OUTROS TRABALHOS DESENVOLVIDOS DURANTE O DOUTORADO

Em conjunto aos capítulos desenvolvidos nesta tese, outros trabalhos e artigos científicos da doutoranda durante este período também focaram em evolução e composição genômica, evolução de elementos repetitivos e suas dinâmicas em vertebrados. Estes renderam artigos em fase de finalização e submissão a periódicos científicos tanto com primeira autoria quanto co-autoria.

5.1 Primeira autoria:

1) Manuscrito submetido para *Zoological Journal of the Linnean Society* (ISSN 1096-3642) (ID ZOJ-01-2024-5693).

"Repetitive DNA genomic analysis of *Leptodactylus pentadactylus* (Amphibia, Anura) reveals low differentiation between sexes and possible incipient evolution of multiple sex chromosomes"

Carolina Crepaldi, Marcelo João da Silva, Thiago Gazoni, Célio Fernando Baptista Haddad, Diogo Cavalcanti Cabral-de-Mello, Patricia Pasquali Parise-Maltempi

2) Manuscrito em fase de finalização para publicação sobre comparação entre mobilomas de macho e fêmea de *Megaleporinus elongatus*, e a relação entre transposons e os satDNAs mais prevalentes no cromossomo heteromórfico sexual da espécie.

3) Manuscrito em fase de finalização para publicação sobre a curadoria e análise do satelitoma da espécie *Leporinus friderici,* com investigação da hipótese da biblioteca entre esta e espécies próximas.

5.2 Co-autoria:

1) Manuscrito submetido para *Cytogenetic and Genome Research* (ISSN 1424-859X)

"Revealing the satellite DNA content in *Ancistrus* sp. (Siluriformes: Loricariidae) by genomic and bioinformatic analysis"

Gabriel Esbrisse dos Santos, **Carolina Crepaldi**, Marcelo João da Silva, Patricia Pasquali Parise-Maltempi

2) Dois manuscritos em fase de elaboração por Murilo Durigon, aluno de Iniciação Científica do laboratório, com satelitomas e análises sobre a hipótese da biblioteca de satDNAs nas espécies de sapo *Thoropa miliaris*, *Leptodactylus pentadactylus* e *Procerathrophrys boiei*.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AHMAD, S. F.; JEHANGIR, M.; SRIKULNATH, K.; MARTINS, C. Fish genomics and its impact on fundamental and applied research of vertebrate biology. **Reviews in Fish Biology and Fisheries**, 2021.

ALMOJIL, D.; BOURGEOIS, Y.; FALIS, M.; HARIYANI, I.; WILCOX, J.; BOISSINOT, S. The Structural, Functional and Evolutionary Impact of Transposable Elements in Eukaryotes. **Genes,** v. 12, n. 6, 2021.

ARKHIPOVA, I.; MESELSON, M. Deleterious transposable elements and the extinction of asexuals. **Bioessays,** v. 27, n. 1, p. 76-85, 2005.

BACHTROG, D. A dynamic view of sex chromosome evolution. **Current opinion in genetics & development,** v. 16, n. 6, p. 578-585, 2006.

BACHTROG, D.; KIRKPATRICK, M.; MANK, J. E.; MCDANIEL, S. F.; PIRES, J. C.; RICE, W.; VALENZUELA, N. Are all sex chromosomes created equal? **Trends in Genetics,** v. 27, n. 9, p. 350-357, 2011.

BACHTROG, D.; MANK, J. E.; PEICHEL, C. L.; KIRKPATRICK, M.; OTTO, S. P.; ASHMAN, T.-L.; HAHN, M. W.; KITANO, J.; MAYROSE, I.; MING, R.; PERRIN, N.; ROSS, L.; VALENZUELA, N.; VAMOSI, J. C.; THE TREE OF SEX, C. Sex Determination: Why So Many Ways of Doing It? **PLOS Biology,** v. 12, n. 7, p. e1001899, 2014.

BELL, E. A.; BUTLER, C. L.; OLIVEIRA, C.; MARBURGER, S.; YANT, L.; TAYLOR, M. I. Transposable element annotation in non-model species: The benefits of species-specific repeat libraries using semi-automated EDTA and DeepTE de novo pipelines. **Molecular Ecology Resources,** v. 22, n. 2, p. 823-833, 2022.

BELYAYEV, A. Bursts of transposable elements as an evolutionary driving force. **Journal of evolutionary biology,** v. 27, n. 12, p. 2573-2584, 2014.

BERGERO, R.; CHARLESWORTH, D. The evolution of restricted recombination in sex chromosomes. **Trends in Ecology & Evolution,** v. 24, n. 2, p. 94-102, 2009.

BIÉMONT, C.; VIEIRA, C. Junk DNA as an evolutionary force. **Nature,** v. 443, n. 7111, p. 521-524, 2006.

BIRINDELLI, J. L. O.; BRITSKI, H. A. Two new species of *Leporinus* (Characiformes: Anostomidae) from the Brazilian Amazon, and redescription of *Leporinus striatus* Kner 1858. **Journal of Fish Biology,** v. 83, n. 5, p. 1128-1160, 2013.

BLANCO, D. R.; VICARI, M. R.; LUI, R. L.; ARTONI, R. F.; DE ALMEIDA, M. C.; TRALDI, J. B.; MARGARIDO, V. P.; MOREIRA-FILHO, O. Origin of the X1X1X2X2/X1X2Y sex chromosome system of *Harttia punctata* (Siluriformes, Loricariidae) inferred from chromosome painting and FISH with ribosomal DNA markers. **Genetica,** v. 142, n. 2, p. 119-126, 2014.

BOURQUE, G.; BURNS, K. H.; GEHRING, M.; GORBUNOVA, V.; SELUANOV, A.; HAMMELL, M.; IMBEAULT, M.; IZSVÁK, Z.; LEVIN, H. L.; MACFARLAN, T. S.; MAGER, D. L.; FESCHOTTE, C. Ten things you should know about transposable elements. **Genome Biology,** v. 19, n. 1, p. 199, 2018.

BUENO, V.; MORESCO, R. M.; KONERAT, J. T.; MOREIRA-FILHO, O.; MARGARIDO, V. P. Instability of Multiple Sex Chromosomes Systems in Fish: The Case of *Erythrinus erythrinus* (Bloch & Schneider, 1801) (Characiformes, Erythrinidae). **Zebrafish,** v. 13, n. 1, p. 26-32, 2015.

CAMACHO; CABRERO, J.; LOPEZ-LEON, M. D.; MARTIN-PECINA, M.; PERFECTTI, F.; GARRIDO-RAMOS, M. A.; RUIZ-RUANO, F. J. On the contingent nature of satellite DNA evolution. **bioRxiv**, 2021a.

CAMACHO; RUIZ-RUANO, F. J.; LÓPEZ-LEÓN, M. D.; CABRERO, J. Satellite DNA Is an Inseparable Fellow Traveler of B Chromosomes. **Progress in molecular and subcellular biology,** v. 60, p. 85-102, 2021b.

CARDUCCI, F.; BARUCCA, M.; CANAPA, A.; BISCOTTI, M. A. Rex retroelements and teleost genomes: an overview. **International journal of molecular sciences,** v. 19, n. 11, p. 3653, 2018.

CARDUCCI, F.; BARUCCA, M.; CANAPA, A.; CAROTTI, E.; BISCOTTI, M. A. Mobile Elements in Ray-Finned Fish Genomes. **Life,** v. 10, n. 10, p. 221, 2020.

CARDUCCI, F.; BISCOTTI, M. A.; BARUCCA, M.; CANAPA, A. Transposable elements in vertebrates: species evolution and environmental adaptation. **The European Zoological Journal,** v. 86, n. 1, p. 497-503, 2019.

CAROTTI, E.; CARDUCCI, F.; CANAPA, A.; BARUCCA, M.; GRECO, S.; GERDOL, M.; BISCOTTI, M. A. Transposable Elements and Teleost Migratory Behaviour. **International Journal of Molecular Sciences,** v. 22, n. 2, 2021.

CHALOPIN, D.; NAVILLE, M.; PLARD, F.; GALIANA, D.; VOLFF, J.-N. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. **Genome Biology and Evolution,** v. 7, n. 2, p. 567-580, 2015a.

CHALOPIN, D.; VOLFF, J.-N.; GALIANA, D.; ANDERSON, J. L.; SCHARTL, M. Transposable elements and early evolution of sex chromosomes in fish. **Chromosome Research,** v. 23, n. 3, p. 545-560, 2015b.

CHARLESWORTH, B. The evolution of sex chromosomes. **Science,** v. 251, n. 4997, p. 1030, 1991.

CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature,** v. 371, n. 6494, p. 215-220, 1994.

CHARLESWORTH, D. Young sex chromosomes in plants and animals. **New Phytologist,** v. 224, n. 3, p. 1095-1107, 2019.

_____. The timing of genetic degeneration of sex chromosomes. **Philosophical Transactions of the Royal Society B: Biological Sciences,** v. 376, n. 1832, p. 20200093, 2021a.

_____. When and how do sex-linked regions become sex chromosomes? **Evolution,** v. 75, n. 3, p. 569-581, 2021b.

CHEN, S.; ZHANG, G.; SHAO, C.; HUANG, Q.; LIU, G.; ZHANG, P.; SONG, W.; AN, N.; CHALOPIN, D.; VOLFF, J.-N. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. **Nature genetics,** v. 46, n. 3, p. 253-260, 2014.

CHEN, S. L.; ZHOU, Q.; SHAO, C. W. Genomic and Epigenetic Aspects of Sex Determination in Half-Smooth Tongue Sole. **Sex Control in Aquaculture**, p. 525-545, 2018.

CIOFFI, M. D. B.; YANO, C. F.; SEMBER, A.; BERTOLLO, L. A. C. Chromosomal Evolution in Lower Vertebrates: Sex Chromosomes in Neotropical Fishes. **Genes,** v. 8, n. 10, p. 258, 2017.

COAN, R. L. B.; MARTINS, C. Landscape of Transposable Elements Focusing on the B Chromosome of the Cichlid Fish *Astatotilapia latifasciata*. **Genes,** v. 9, n. 6, p. 269, 2018.

COLONNA ROMANO, N.; FANTI, L. Transposable Elements: Major Players in Shaping Genomic and Evolutionary Patterns. **Cells,** v. 11, n. 6, p. 1048, 2022.

CREPALDI, C.; MARTÍ, E.; GONÇALVES, É. M.; MARTÍ, D. A.; PARISE-MALTEMPI, P. P. Genomic differences between the sexes in a fish species seen through satellite DNAs. **Frontiers in Genetics,** v. 12, 2021.

CREPALDI, C.; PARISE-MALTEMPI, P. P. Heteromorphic Sex Chromosomes and Their DNA Content in Fish: An Insight through Satellite DNA Accumulation in *Megaleporinus elongatus*. **Cytogenetic and Genome Research,** v. 160, n. 1, p. 38-46, 2020.

DA SILVA, E. L.; DE BORBA, R. S.; PARISE-MALTEMPI, P. P. Chromosome mapping of repetitive sequences in Anostomidae species: implications for genomic and sex chromosome evolution. **Molecular Cytogenetics,** v. 5, n. 1, p. 45, 2012.

DE BARROS, L. C.; PISCOR, D.; PARISE-MALTEMPI, P. P.; FELDBERG, E. Differentiation and Evolution of the W Chromosome in the Fish Species of *Megaleporinus* (Characiformes, Anostomidae). **Sexual Development,** v. 12, n. 4, p. 204-209, 2018.

DE FREITAS, N. L.; AL-RIKABI, A. B.; BERTOLLO, L. A. C.; EZAZ, T.; YANO, C. F.; DE OLIVEIRA, E. A.; HATANAKA, T.; DE BELLO CIOFFI, M. Early stages of XY sex chromosomes differentiation in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae) revealed by DNA repeats accumulation. **Current genomics,** v. 19, n. 3, p. 216-226, 2018.

DE OLIVEIRA, E. A.; SEMBER, A.; BERTOLLO, L. A. C.; YANO, C. F.; EZAZ, T.; MOREIRA-FILHO, O.; HATANAKA, T.; TRIFONOV, V.; LIEHR, T.; AL-RIKABI, A. B. H.; RÁB, P.; PAINS, H.; CIOFFI, M. D. B. Tracking the evolutionary pathway of sex chromosomes among fishes: characterizing the unique XX/XY1Y2 system in *Hoplias malabaricus* (Teleostei, Characiformes). **Chromosoma,** v. 127, n. 1, p. 115-128, 2018.

DECHAUD, C.; VOLFF, J.-N.; SCHARTL, M.; NAVILLE, M. Sex and the TEs: transposable elements in sexual development and function in animals. **Mobile DNA,** v. 10, n. 1, p. 42, 2019.

DESJARDINS, J.; FERNALD, R. Fish sex: why so diverse? **Current opinion in neurobiology,** v. 19, n. 6, p. 648, 2009.

DOS SANTOS, R. Z.; CALEGARI, R. M.; SILVA, D. M. Z. D. A.; RUIZ-RUANO, F. J.; MELO, S.; OLIVEIRA, C.; FORESTI, F.; ULIANO-SILVA, M.; PORTO-FORESTI, F.; UTSUNOMIA, R. A Long-Term Conserved Satellite DNA That Remains Unexpanded in Several Genomes of Characiformes Fish Is Actively Transcribed. **Genome Biology and Evolution,** v. 13, n. 2, 2021.

DOVER, G. A. Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. **Trends in Genetics,** v. 2, p. 159-165, 1986.

DOVER, G. A.; LINARES, A. R.; BOWEN, T.; HANCOCK, J. M. [38] Detection and quantification of concerted evolution and molecular drive. **Methods in enzymology,** v. 224, p. 525-541, 1993.

DRUMMOND, A.; ASHTON, B.; CHEUNG, M.; HELED, J.; KEARSE, M.; MOIR, R.; STONES-HAVAS, S.; THIERER, T.; WILSON, A. Geneious v4. 8. 2009. **DOI: http://www. geneious. com. Accessed,** v. 21, 2015.

DULZ, T. A.; AZAMBUJA, M.; LORSCHEIDER, C. A.; NOLETO, R. B.; MOREIRA-FILHO, O.; NOGAROTO, V.; NASCIMENTO, V. D.; DINIZ, D.; DE MELLO AFFONSO, P. R. A.; VICARI, M. R. Repetitive DNAs and chromosome evolution in Megaleporinus obtusidens and M. reinhardti (Characiformes: Anostomidae). **Genetica**, 2024.

DULZ, T. A.; LORSCHEIDER, C. A.; NASCIMENTO, V. D.; NOLETO, R. B.; MOREIRA-FILHO, O.; NOGAROTO, V.; VICARI, M. R. Comparative cytogenetics among *Leporinus friderici* and *Leporellus vittatus* populations (Characiformes, Anostomidae): focus on repetitive DNA elements. **Comparative cytogenetics,** v. 13, n. 2, p. 1, 2019.

EDGAR, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics,** v. 5, n. 1, p. 113, 2004.

ELLEGREN, H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. **Nature Reviews Genetics,** v. 12, n. 3, p. 157-166, 2011.

ESCHMEYER, W.; FONG, J. **Catalog of fishes: genera, species, references. San Francisco: California Academy of Sciences** 2015.

ETCHEGARAY, E.; DECHAUD, C.; BARBIER, J.; NAVILLE, M.; VOLFF, J.-N. Diversity of Harbinger-like Transposons in Teleost Fish Genomes. **Animals,** v. 12, n. 11, p. 1429, 2022.

FERREIRA, D. C.; PORTO-FORESTI, F.; OLIVEIRA, C.; FORESTI, F. Transposable elements as a potential source for understanding the fish genome. **Mobile Genetic Elements,** v. 1, n. 2, p. 112-117, 2011.

FESCHOTTE, C.; PRITHAM, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. **Annual Review of Genetics,** v. 41, n. 1, p. 331-368, 2007.

FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research,** v. 39, n. suppl_2, p. W29-W37, 2011.

FLUTRE, T.; DUPRAT, E.; FEUILLET, C.; QUESNEVILLE, H. Considering Transposable Element Diversification in De Novo Annotation Approaches. **PLOS ONE,** v. 6, n. 1, p. e16526, 2011.

FRY, K.; SALSER, W. Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and characterization of similar sequences in other rodents. **Cell,** v. 12, n. 4, p. 1069-1084, 1977.

FU, L.; NIU, B.; ZHU, Z.; WU, S.; LI, W. CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics,** v. 28, n. 23, p. 3150-3152, 2012.

FURMAN, B. L.; METZGER, D. C.; DAROLTI, I.; WRIGHT, A. E.; SANDKAM, B. A.; ALMEIDA, P.; SHU, J. J.; MANK, J. E. Sex chromosome evolution: So many exceptions to the rules. **Genome Biology and Evolution,** v. 12, n. 6, p. 750-763, 2020.

GALETTI JR, P.; FORESTI, F. Evolution of the ZZ/ZW system in *Leporinus* (Pisces, Anostomidae). **Cytogenetic and Genome Research,** v. 43, n. 1-2, p. 43-46, 1986.

GAMMERDINGER, W. J.; KOCHER, T. D. Unusual Diversity of Sex Chromosomes in African Cichlid Fishes. **Genes,** v. 9, n. 10, p. 480, 2018.

GAO, B.; SHEN, D.; XUE, S.; CHEN, C.; CUI, H.; SONG, C. The contribution of transposable elements to size variations between four teleost genomes. **Mobile DNA,** v. 7, n. 1, p. 4, 2016.

GARRIDO-RAMOS, M. A. Satellite DNA: An Evolving Topic. **Genes,** v. 8, n. 9, 2017.

GOES, C. A. G.; DOS SANTOS, N.; RODRIGUES, P. H. D. M.; STORNIOLI, J. H. F.; SILVA, A. B. D.; DOS SANTOS, R. Z.; VIDAL, J. A. D.; SILVA, D. M. Z. D. A.; ARTONI, R. F.; FORESTI, F.; HASHIMOTO, D. T.; PORTO-FORESTI, F.; UTSUNOMIA, R. The Satellite DNA Catalogues of Two Serrasalmidae (Teleostei, Characiformes): Conservation of General satDNA Features over 30 Million Years. **Genes,** v. 14, n. 1, p. 91, 2023.

GRAVES, J. A. M. Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. **Annual Review of Genetics,** v. 42, n. 1, p. 565-586, 2008.

HASHIMOTO, D. T.; PARISE-MALTEMPI, P. P.; LAUDICINA, A.; BORTOLOZZI, J.; SENHORINI, J. A.; FORESTI, F.; PORTO-FORESTI, F. Repetitive DNA probe linked to sex chromosomes in hybrids between Neotropical fish *Leporinus macrocephalus* and *Leporinus elongatus* (Characiformes, Anostomidae). **Cytogenetic and Genome Research,** v. 124, n. 2, p. 151-157, 2009.

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

JAGANNATHAN, M.; CUMMINGS, R.; YAMASHITA, Y. M. A conserved function for pericentromeric satellite DNA. **Elife,** v. 7, p. e34122, 2018.

JURKA, J.; KAPITONOV, V. V.; PAVLICEK, A.; KLONOWSKI, P.; KOHANY, O.; WALICHIEWICZ, J. Repbase Update, a database of eukaryotic repetitive elements. **Cytogenetic and Genome Research,** v. 110, n. 1-4, p. 462-467, 2005.

KEJNOVSKY, E.; HOBZA, R.; CERMAK, T.; KUBAT, Z.; VYSKOT, B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. **Heredity,** v. 102, p. 533, 2009.

KENT, T. V.; UZUNOVIĆ, J.; WRIGHT, S. I. Coevolution between transposable elements and recombination. **Philosophical Transactions of the Royal Society B: Biological Sciences,** v. 372, n. 1736, p. 20160458, 2017.

KITANO, J.; PEICHEL, C. L. Turnover of sex chromosomes and speciation in fishes. **Environmental Biology of Fishes,** v. 94, n. 3, p. 549-558, 2012.

KOHANY, O.; GENTLES, A. J.; HANKUS, L.; JURKA, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **BMC Bioinformatics,** v. 7, n. 1, p. 474, 2006.

KRATOCHVÍL, L.; STÖCK, M.; ROVATSOS, M.; BULLEJOS, M.; HERPIN, A.; JEFFRIES, D. L.; PEICHEL, C. L.; PERRIN, N.; VALENZUELA, N.; POKORNÁ, M. J. Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. **Philosophical Transactions of the Royal Society B,** v. 376, n. 1833, p. 20200097, 2021.

LEHMANN, R.; KOVAŘÍK, A.; OCALEWICZ, K.; KIRTIKLIS, L.; ZUCCOLO, A.; TEGNER, J. N.; WANZENBÖCK, J.; BERNATCHEZ, L.; LAMATSCH, D. K.; SYMONOVÁ, R. DNA Transposon Expansion is Associated with Genome Size Increase in Mudminnows. **Genome Biology and Evolution,** v. 13, n. 10, p. evab228, 2021.

LI, S.-F.; GUO, Y.-J.; LI, J.-R.; ZHANG, D.-X.; WANG, B.-X.; LI, N.; DENG, C.-L.; GAO, W.-J. The landscape of transposable elements and satellite DNAs in the genome of a dioecious plant spinach (*Spinacia oleracea* L.). **Mobile DNA,** v. 10, n. 1, p. 3, 2019.

LIAO, X.; ZHU, W.; ZHOU, J.; LI, H.; XU, X.; ZHANG, B.; GAO, X. Repetitive DNA sequence detection and its role in the human genome. **Communications Biology,** v. 6, n. 1, p. 954, 2023.

LORITE, P.; CARRILLO, J. A.; TINAUT, A.; PALOMEQUE, T. Evolutionary dynamics of satellite DNA in species of the Genus *Formica* (Hymenoptera, Formicidae). **Gene,** v. 332, p. 159-168, 2004.

LORITE, P.; MUÑOZ-LÓPEZ, M.; CARRILLO, J. A.; SANLLORENTE, O.; VELA, J.; MORA, P.; TINAUT, A.; TORRES, M. I.; PALOMEQUE, T. Concerted evolution, a slow process for ant satellite DNA: study of the satellite DNA in the Aphaenogaster genus (Hymenoptera, Formicidae). **Organisms Diversity & Evolution,** v. 17, n. 3, p. 595-606, 2017.

MAKAŁOWSKI, W. Genomic scrap yard: how genomes utilize all that junk. **Gene,** v. 259, n. 1, p. 61-67, 2000.

MAKAŁOWSKI, W.; GOTEA, V.; PANDE, A.; MAKAŁOWSKA, I. Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. In: ANISIMOVA, M. (Ed.). **Evolutionary Genomics: Statistical and Computational Methods**. New York, NY: Springer New York, 2019. p.177-207. ISBN 978-1-4939-9074-0.

MANK, J. E.; AVISE, J. C. Evolutionary Diversity and Turn-Over of Sex Determination in Teleost Fishes. **Sexual Development,** v. 3, n. 2-3, p. 60-67, 2009.

MANK, J. E.; PROMISLOW, D. E.; AVISE, J. C. Evolution of alternative sex-determining mechanisms in teleost fishes. **Biological Journal of the Linnean Society,** v. 87, n. 1, p. 83-93, 2006.

MARRETA, M. E.; FALDONI, F. L. C.; PARISE-MALTEMPI, P. P. Cytogenetic mapping of the W chromosome in the genus *Leporinus* (Teleostei, Anostomidae) using a highly repetitive DNA sequence. **Journal of Fish Biology,** v. 80, n. 3, p. 630-637, 2012.

MARTINS, C.; GALETTI, P. M. Chromosomal Localization of 5S rDNA Genes in Leporinus Fish (Anostomidae, Characiformes). **Chromosome Research,** v. 7, n. 5, p. 363-367, 1999.

MÉROT, C.; OOMEN, R. A.; TIGANO, A.; WELLENREUTHER, M. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. **Trends in Ecology & Evolution,** v. 35, n. 7, p. 561-572, 2020.

MEŠTROVIĆ, N.; CASTAGNONE-SERENO, P.; PLOHL, M. Interplay of Selective Pressure and Stochastic Events Directs Evolution of the MEL172 Satellite DNA Library in Root-Knot Nematodes. **Molecular Biology and Evolution,** v. 23, n. 12, p. 2316-2325, 2006.

MEŠTROVIĆ, N.; MRAVINAC, B.; PAVLEK, M.; VOJVODA-ZELJKO, T.; ŠATOVIĆ, E.; PLOHL, M. Structural and functional liaisons between transposable elements and satellite DNAs. **Chromosome research,** v. 23, n. 3, p. 583-596, 2015.

MISTRY, J.; CHUGURANSKY, S.; WILLIAMS, L.; QURESHI, M.; SALAZAR, G. A.; SONNHAMMER, E. L.; TOSATTO, S. C.; PALADIN, L.; RAJ, S.; RICHARDSON, L. J. Pfam: The protein families database in 2021. **Nucleic acids research,** v. 49, n. D1, p. D412-D419, 2021.

MYOSHO, T.; TAKEHANA, Y.; HAMAGUCHI, S.; SAKAIZUMI, M. Turnover of sex chromosomes in celebensis group medaka fishes. **G3: Genes, Genomes, Genetics,** v. 5, n. 12, p. 2685-2691, 2015.

NAKAYAMA, I.; FORESTI, F.; TEWARI, R.; SCHARTL, M.; CHOURROUT, D. Sex chromosome polymorphism and heterogametic males revealed by two cloned DNA probes in the ZW/ZZ fish *Leporinus elongatus*. **Chromosoma,** v. 103, n. 1, p. 31-39, 1994.

NOVÁK, P.; NEUMANN, P.; MACAS, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. **Nature protocols,** v. 15, n. 11, p. 3745-3776, 2020.

NOVÁK, P.; NEUMANN, P.; PECH, J.; STEINHAISL, J.; MACAS, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics,** v. 29, n. 6, p. 792-793, 2013.

O'NEILL, K.; BROCKS, D.; HAMMELL, M. G. Mobile genomics: tools and techniques for tackling transposons. **Philosophical Transactions of the Royal Society B,** v. 375, n. 1795, p. 20190345, 2020.

PAÇO, A.; FREITAS, R.; VIEIRA-DA-SILVA, A. Conversion of DNA sequences: From a transposable element to a tandem repeat or to a gene. **Genes,** v. 10, n. 12, p. 1014, 2019.

PALMER, D. H.; ROGERS, T. F.; DEAN, R.; WRIGHT, A. E. How to identify sex chromosomes and their turnover. **Molecular ecology**, 2019.

PARISE-MALTEMPI, P. P.; DA SILVA, E. L.; RENS, W.; DEARDEN, F.; O'BRIEN, P. C.; TRIFONOV, V.; FERGUSON-SMITH, M. A. Comparative analysis of sex chromosomes in *Leporinus* species (Teleostei, Characiformes) using chromosome painting. **BMC Genetics,** v. 14, n. 1, p. 60, 2013.

PARISE-MALTEMPI, P. P.; MARTINS, C.; OLIVEIRA, C.; FORESTI, F. Identification of a new repetitive element in the sex chromosomes of *Leporinus elongatus* (Teleostei: Characiformes: Anostomidae): new insights into the sex chromosomes of *Leporinus*. **Cytogenetic and Genome Research,** v. 116, n. 3, p. 218-223, 2007.

PEZER, Ž.; BRAJKOVIĆ, J.; FELICIELLO, I.; UGARKOVIĆ, Đ. Satellite DNA-mediated effects on genome regulation. In: (Ed.). **Repetitive DNA**: Karger Publishers, v.7, 2012. p.153-169.

PLOHL; MESTROVIC, N.; MRAVINAC, B. Satellite DNA evolution. **Genome Dyn,** v. 7, p. 126-52, 2012a.

PLOHL, M.; LUCHETTI, A.; MESTROVIC, N.; MANTOVANI, B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. **Gene,** v. 409, n. 1-2, p. 72-82, 2008.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA Evolution. v. - 7, p. - 152, 2012b.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Centromere identity from the DNA point of view. **Chromosoma,** v. 123, n. 4, p. 313-325, 2014.

POLTRONIERI, J.; MARQUIONI, V.; BERTOLLO, L. A. C.; KEJNOVSKY, E.; MOLINA, W. F.; LIEHR, T.; CIOFFI, M. B. Comparative Chromosomal Mapping of Microsatellites in *Leporinus* Species (Characiformes, Anostomidae): Unequal Accumulation on the W Chromosomes. **Cytogenetic and Genome Research,** v. 142, n. 1, p. 40-45, 2014.

POTTER, S. C.; LUCIANI, A.; EDDY, S. R.; PARK, Y.; LOPEZ, R.; FINN, R. D. HMMER web server: 2018 update. **Nucleic Acids Research,** v. 46, n. W1, p. W200-W204, 2018.

R CORE TEAM. **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing 2021.

RAMIREZ, J. L.; BIRINDELLI, J. L. O.; GALETTI, P. M. A new genus of Anostomidae (Ostariophysi: Characiformes): Diversity, phylogeny and biogeography based on cytogenetic, molecular and morphological data. **Molecular Phylogenetics and Evolution,** v. 107, p. 308-323, 2017a.

RAMIREZ, J. L.; BIRINDELLI, J. L. O.; GALETTI, P. M., JR. A new genus of Anostomidae (Ostariophysi: Characiformes): Diversity, phylogeny and biogeography based on cytogenetic, molecular and morphological data. **Mol Phylogenet Evol,** v. 107, p. 308-323, 2017b.

RENAUT, S.; BERNATCHEZ, L. Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus spp*. Salmonidae). **Heredity,** v. 106, n. 6, p. 1003-1011, 2011.

RODRIGUES, P. H. D. M.; DOS SANTOS, R. Z.; SILVA, D. M. Z. D. A.; GOES, C. A. G.; OLIVEIRA, C.; FORESTI, F.; PORTO-FORESTI, F.; UTSUNOMIA, R. Chromosomal and genomic dynamics of satellite DNAs in Characidae (Characiformes, Teleostei) Species. **Zebrafish,** v. 16, n. 4, p. 408-414, 2019.

RODRÍGUEZ, M. E.; CROSS, I.; ARIAS-PÉREZ, A.; PORTELA-BENS, S.; MERLO, M. A.; LIEHR, T.; REBORDINOS, L. Cytogenomics Unveil Possible Transposable Elements Driving Rearrangements in Chromosomes 2 and 4 of *Solea senegalensis*. **International Journal of Molecular Sciences,** v. 22, n. 4, p. 1614, 2021.

ROGERS, R. L.; ZHOU, L.; CHU, C.; MÁRQUEZ, R.; CORL, A.; LINDEROTH, T.; FREEBORN, L.; MACMANES, M. D.; XIONG, Z.; ZHENG, J.; GUO, C.; XUN, X.; KRONFORST, M. R.; SUMMERS, K.; WU, Y.; YANG, H.; RICHARDS-ZAWACKI, C. L.; ZHANG, G.; NIELSEN, R. Genomic Takeover by Transposable Elements in the Strawberry Poison Frog. **Molecular Biology and Evolution,** v. 35, n. 12, p. 2913-2927, 2018.

ROSOLEN, L. A. M.; VICARI, M. R.; ALMEIDA, M. C. Accumulation of Transposable Elements in Autosomes and Giant Sex Chromosomes of *Omophoita* (Chrysomelidae: Alticinae). **Cytogenetic and Genome Research,** v. 156, n. 4, p. 215-222, 2018.

RUIZ-RUANO, F. J.; CABRERO, J.; LÓPEZ-LEÓN, M. D.; CAMACHO, J. P. M. Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. **Chromosoma,** v. 126, n. 4, p. 487-500, 2017.

RUIZ-RUANO, F. J.; LÓPEZ-LEÓN, M. D.; CABRERO, J.; CAMACHO, J. P. M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports,** v. 6, p. 28333, 2016.

ŠATOVIĆ-VUKŠIĆ, E.; PLOHL, M. Satellite DNA - From Localized to Highly Dispersed Genome Components. **Genes,** v. 14, n. 3, p. 742, 2023.

SCHARTL, M.; SCHMID, M.; NANDA, I. Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. **Chromosoma,** v. 125, n. 3, p. 553-571, 2016.

SCHEMBERGER, M. O.; BELLAFRONTE, E.; NOGAROTO, V.; ALMEIDA, M. C.; SCHÜHLI, G. S.; ARTONI, R. F.; MOREIRA-FILHO, O.; VICARI, M. R. Differentiation of repetitive DNA sites and sex chromosome systems reveal closely related group in Parodontidae (Actinopterygii: Characiformes). **Genetica,** v. 139, n. 11, p. 1499-1508, 2011.

SCHEMBERGER, M. O.; NASCIMENTO, V. D.; COAN, R.; RAMOS, É.; NOGAROTO, V.; ZIEMNICZAK, K.; VALENTE, G. T.; MOREIRA-FILHO, O.; MARTINS, C.; VICARI, M. R. DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. **Chromosoma,** v. 128, n. 4, p. 547-560, 2019.

SCHMIEDER, R.; EDWARDS, R. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. **PLOS ONE,** v. 6, n. 3, p. e17288, 2011.

SEMBER, A.; BERTOLLO, L. A. C.; RAB, P.; YANO, C. F.; HATANAKA, T.; DE OLIVEIRA, E. A.; CIOFFI, M. B. Sex Chromosome Evolution and Genomic Divergence in the Fish *Hoplias malabaricus* (Characiformes, Erythrinidae). **Front Genet,** v. 9, p. 71, 2018.

SEMBER, A.; NGUYEN, P.; PEREZ, M. F.; ALTMANOVÁ, M.; RÁB, P.; CIOFFI, M. D. B. Multiple sex chromosomes in teleost fishes from a cytogenetic perspective: state of the art and future challenges. **Philosophical Transactions of the Royal Society B,** v. 376, n. 1833, p. 20200098, 2021.

SERRATO-CAPUCHINA, A.; MATUTE, D. R. The Role of Transposable Elements in Speciation. **Genes,** v. 9, n. 5, p. 254, 2018.

SHAO, F.; HAN, M.; PENG, Z. Evolution and diversity of transposable elements in fish genomes. **Scientific Reports,** v. 9, n. 1, p. 15399, 2019.

SILVA, D. M. Z. D. A.; UTSUNOMIA, R.; RUIZ-RUANO, F. J.; DANIEL, S. N.; PORTO-FORESTI, F.; HASHIMOTO, D. T.; OLIVEIRA, C.; CAMACHO, J. P. M.; FORESTI, F. High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. **Scientific Reports,** v. 7, n. 1, p. 12726, 2017.

SMIT, A.; HUBLEY, R.; GREEN, P. **RepeatMasker Open v. 4.0. 6** 2017.

SOTERO-CAIO, C. G.; PLATT, R. N., II; SUH, A.; RAY, D. A. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. **Genome Biology and Evolution,** v. 9, n. 1, p. 161-177, 2017.

SPLENDORE DE BORBA, R.; LOURENÇO DA SILVA, E.; PARISE-MALTEMPI, P. P. Chromosome mapping of retrotransposable elements *Rex1* and *Rex3* in *Leporinus* Spix, 1829 species (Characiformes: Anostomidae) and its relationships among heterochromatic segments and W sex chromosome. **Mobile Genetic Elements,** v. 3, n. 6, p. e27460, 2013.

SRIKULNATH, K.; AHMAD, S. F.; SINGCHAT, W.; PANTHUM, T. Do Ty3/Gypsy Transposable Elements Play Preferential Roles in Sex Chromosome Differentiation? **Life,** v. 12, n. 4, p. 522, 2022.

STORER, J.; HUBLEY, R.; ROSEN, J.; WHEELER, T. J.; SMIT, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. **Mobile DNA,** v. 12, n. 1, p. 1-14, 2021.

STORNIOLI, J. H. F.; GOES, C. A. G.; CALEGARI, R. M.; DOS SANTOS, R. Z.; GIGLIO, L. M.; FORESTI, F.; OLIVEIRA, C.; PENITENTE, M.; PORTO-FORESTI, F.; UTSUNOMIA, R. The B Chromosomes of *Prochilodus lineatus* (Teleostei, Characiformes) Are Highly Enriched in Satellite DNAs. **Cells,** v. 10, n. 6, p. 1527, 2021.

SUNTRONPONG, A.; SINGCHAT, W.; KRUASUWAN, W.; PRAKHONGCHEEP, O.; SILLAPAPRAYOON, S.; MUANGMAI, N.; SOMYONG, S.; INDANADA, C.; KRAICHAK, E.; PEYACHOKNAGUL, S. Characterization of centromeric satellite DNAs (MALREP) in the Asian swamp eel (*Monopterus albus*) suggests the possible origin of repeats from transposable elements. **Genomics**, 2020.

THAKUR, J.; PACKIARAJ, J.; HENIKOFF, S. Sequence, Chromatin and Evolution of Satellite DNA. **International Journal of Molecular Sciences,** v. 22, n. 9, 2021.

UGARKOVIĆ, Đ. **Satellite DNAs in Physiology and Evolution**: Springer 2021.

UGARKOVIĆ, Đ.; PLOHL, M. Variation in satellite DNA profiles—causes and effects. **The EMBO Journal,** v. 21, n. 22, p. 5955, 2002.

UTSUNOMIA, R.; RUIZ-RUANO, F. J.; SILVA, D.; SERRANO, E. A.; ROSA, I. F.; SCUDELER, P. E. S.; HASHIMOTO, D. T.; OLIVEIRA, C.; CAMACHO, J. P. M.; FORESTI, F. A Glimpse into the Satellite DNA Library in Characidae Fish (Teleostei, Characiformes). **Front Genet,** v. 8, p. 103, 2017.

UTSUNOMIA, R.; SILVA, D. M. Z. D. A.; RUIZ-RUANO, F. J.; GOES, C. A. G.; MELO, S.; RAMOS, L. P.; OLIVEIRA, C.; PORTO-FORESTI, F.; FORESTI, F.; HASHIMOTO, D. T. Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. **Scientific Reports,** v. 9, n. 1, p. 5856, 2019.

VALENTE, G. T.; MAZZUCHELLI, J.; FERREIRA, I. A.; POLETTO, A. B.; FANTINATTI, B. E. A.; MARTINS, C. Cytogenetic Mapping of the Retroelements Rex1, Rex3 and Rex6 among Cichlid Fish: New

Insights on the Chromosomal Distribution of Transposable Elements. **Cytogenetic and Genome Research,** v. 133, n. 1, p. 34-42, 2010.

VENERE, P. C.; ALVES FERREIRA, I.; MARTINS, C.; MANOEL GALETTI, P. A Novel ZZ/ZW Sex Chromosome System for the Genus *Leporinus* (Pisces, Anostomidae, Characiformes). **Genetica,** v. 121, n. 1, p. 75-80, 2004.

VICARI, M. R.; NOGAROTO, V.; NOLETO, R. B.; CESTARI, M. M.; CIOFFI, M. B.; ALMEIDA, M. C.; MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; ARTONI, R. F. Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. **Journal of Fish Biology,** v. 76, n. 5, p. 1094-1116, 2010.

VOLFF, J. N. Genome evolution and biodiversity in teleost fish. **Heredity,** v. 94, n. 3, p. 280-294, 2005.

WANG, Y.; WANG, Q.; HUANG, H.; HUANG, W.; CHEN, Y.; MCGARVEY, P. B.; WU, C. H.; ARIGHI, C. N.; CONSORTIUM, U. A crowdsourcing open platform for literature curation in UniProt. **PLoS biology,** v. 19, n. 12, p. e3001464, 2021.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A. H. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics,** v. 8, n. 12, p. 973-982, 2007.

WRIGHT, A. E.; DEAN, R.; ZIMMER, F.; MANK, J. E. How to make a sex chromosome. **Nature Communications,** v. 7, p. 12087, 2016.

YANO; BERTOLLO, L. A. C.; EZAZ, T.; TRIFONOV, V.; SEMBER, A.; LIEHR, T.; CIOFFI, M. B. Highly conserved Z and molecularly diverged W chromosomes in the fish genus *Triportheus* (Characiformes, Triportheidae). **Heredity,** v. 118, p. 276, 2016a.

YANO, C. F.; BERTOLLO, L. A. C.; LIEHR, T.; TROY, W. P.; CIOFFI, M. D. B. W Chromosome Dynamics in *Triportheus* Species (Characiformes, Triportheidae): An Ongoing Process Narrated by Repetitive Sequences. **Journal of Heredity,** v. 107, n. 4, p. 342-348, 2016b.

YOSHIDA, K.; MAKINO, T.; YAMAGUCHI, K.; SHIGENOBU, S.; HASEBE, M.; KAWATA, M.; KUME, M.; MORI, S.; PEICHEL, C. L.; TOYODA, A. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. **PLoS genetics,** v. 10, n. 3, p. e1004223, 2014.

YUAN, Z.; LIU, S.; ZHOU, T.; TIAN, C.; BAO, L.; DUNHAM, R.; LIU, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. **BMC Genomics,** v. 19, n. 1, p. 141, 2018.