


---

**CIÊNCIAS BIOLÓGICAS**

---

**ÉVELIN MARIANI GONÇALVES**

**ANÁLISE COMPARATIVA DE UM DNA SATÉLITE  
COMPARTILHADO EM ESPÉCIES DE PEIXES  
CHARACIFORMES**



Rio Claro - SP  
2022

ÉVELIN MARIANI GONÇALVES

**ANÁLISE COMPARATIVA DE UM DNA SATÉLITE COMPARTILHADO  
EM ESPÉCIES DE PEIXES CHARACIFORMES**

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências – Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do grau de Bacharela e Licenciada em Ciências Biológicas.

Orientadora: Profa. Dra. Patrícia Pasquali Parise-Maltempi

Co-orientadora: Ma. Carolina Crepaldi

Rio Claro - SP  
2022

G635a      Gonçalves, Évelin Mariani  
Análise comparativa de um DNA satélite compartilhado em espécies  
de peixes Characiformes / Évelin Mariani Gonçalves. -- Rio Claro,  
2022  
46 p.

Trabalho de conclusão de curso (Bacharelado e licenciatura -  
Ciências Biológicas) - Universidade Estadual Paulista (Unesp),  
Instituto de Biociências, Rio Claro  
Orientadora: Patrícia Pasquali Parise-Maltempi  
Coorientadora: Carolina Crepaldi

1. Biologia molecular. 2. DNA. 3. Characiformes. I. Título.

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

Essa ficha não pode ser modificada.

ÉVELIN MARIANI GONÇALVES

**ANÁLISE COMPARATIVA DE UM DNA SATÉLITE COMPARTILHADO  
EM ESPÉCIES DE PEIXES CHARACIFORMES**

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências – Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do grau de Bacharela e Licenciada em Ciências Biológicas.


BANCA EXAMINADORA:

Profa. Dra. Patrícia Pasquali Parise-Maltempi (orientadora)  
Ma. Carolina Crepaldi (co-orientadora)  
Me. Marcelo João da Silva  
Dra. Luiza Rieder Cholak

Aprovado em: 07 de Janeiro de 2022



Assinatura do discente



Assinatura do(a) orientador(a)



Assinatura do(a) coorientador(a)

Dedico este trabalho a todas as pessoas que consagram suas vidas à pesquisa brasileira e que resistem com veemência aos tempos de negação. A vocês todo o meu respeito.

## AGRADECIMENTOS

Agradeço, em primeiro lugar, às grandes mulheres da minha vida, mãe Luciana, bisá Ernestina, vó Maria e tia Le, vocês me ensinaram a ter a garra da mulher pobre e nordestina, obrigada pela força, por resistirem ao meu lado e pelo amor incondicional! Mãezinha, palavras não conseguem mensurar o quanto sou profundamente grata a você. Obrigada pela amizade, pelo cuidado e apoio, por me doar tanto amor e sempre se orgulhar de mim, se cheguei até aqui foi porque a senhora sempre acreditou que eu conseguiria, é uma felicidade imensa ser a sua filha. Te amo mais que tudo! Um beijo por um queijo!

À minha irmã Letícia, que sempre me conforta com os melhores conselhos, cafunés e massagem nos pés. Obrigada pelo companheirismo, fidelidade e por comemorar todas minhas conquistas (mesmo sem entender nada), ser a sua irmã é uma eterna aventura! Te amo pra sempre, Letucinha.

Aos meus primos Vinicius, Vitor, Igor, Karla e Jeff, vocês são a melhor família que eu poderia ter. Agradeço especialmente a quem sempre me encoraja, meu primo Vinicius, obrigada pelo carinho, por acreditar em mim e por todas as nossas tardes de cantoria.

Ao meu pai Roberto, que dedicou parte de sua vida para me oferecer um ensino de qualidade. Agradeço pelos preciosos momentos de lazer, que me trouxeram muita conversa solta, risadas e um bocado de cerveja boa.

À minha orientadora Patrícia P. Parise-Maltempí, que me acolheu nos meus anos iniciais e me mostrou com muita paixão o quanto é lindo o mundo dos cromossomos. Obrigada por toda orientação, pela motivação, pela confiança, por me abrir grandes caminhos e por sempre lembrar o quanto sou capaz. Serei eternamente grata pela oportunidade e pelo carinho que você me entregou durante esses 4 anos. Meus mais sinceros agradecimentos.

À Carolina Crepaldi, que além de co-orientadora, se tornou uma grande amiga. Agradeço pela paciência, carinho e cuidado durante todos esses anos. Você me fez amar a ciência (até nos surtos), a progredir como pesquisadora e como pessoa. Obrigada por ouvir minhas aflições, por acolher e resolver todos os meus problemas com tamanha tranquilidade e ternura, você é a melhor amiga e companheira de pesquisa que alguém poderia ter. Foi uma experiência maravilhosa ser sua pupila. Deixo aqui minha eterna admiração.

Aos meus amigos de laboratório, Pati, Marcelo, Murilo, Raquel, Luiza, Gabriel, Giselle, Luciana, Matheus e Carol, tamanha é a minha felicidade de ter compartilhado essa bancada com vocês. Obrigada pelos cafés e pelas comemorações acompanhadas de ótimas comidas. Agradeço principalmente ao Marcelo, Murilo, Raquel e Carol, vocês me abraçaram e trouxeram a calma nos dias atribulados, obrigada pela amizade!

Aos meus professores, Guima e Ruy, sempre lembrarei de vocês com muito carinho e reconhecimento por tudo o que vocês fizeram para que eu pudesse entrar e permanecer na Unesp. Também agradeço aos professores Roberto Goiten, Luiz Felipe Daibes e Denis Peixoto, que foram imprescindíveis para a minha formação, vocês realizaram um ensino humanizador, com muito amor, empatia e acolhimento. Obrigada por me fazerem enxergar além da biologia.

Aos meus grandes e queridos amigos, Mió e Pe, que estão sempre dispostos a me ouvir e aconselhar. Obrigada por aparecerem na fase mais louca da minha vida e por permanecerem, me trazendo tanta luz e carinho. Mil vezes obrigada!

Aos meus amigos da Unesp, seria impossível elencar cada pessoa que me marcou e me ajudou de alguma forma, mas saibam que vocês foram essenciais nessa minha jornada. Agradeço por partilharem comigo histórias e experiências de um jeito que só um louco unespiano saberia fazer. Realmente, quem é de Rio Claro é mais feliz!

Ao Movimento Estudantil, por me ensinar que a Universidade se faz nas salas, nas extensões, nas pesquisas e principalmente na luta. Vocês me ensinaram a ressignificar os espaços, a criar coragem e a lutar pelos meus ideais. Eu conheci cada pedacinho dessa Universidade, resisti, desbravei e lutei por mim, por vocês e por aqueles que virão. Minha luta fica aqui para que eu não seja reconhecida como a única da família a concluir uma Universidade pública, mas para que eu seja lembrada como a primeira de muitas.

Ao agregado mais querido da República Pitayas, Rodrigo (ou Babylookson para os íntimos), que sempre foi tão atencioso e me entregou ótimos conselhos, obrigada, amigo, pelos cafés e pelos pratos deliciosos que você sempre faz (eles salvaram meus semestres)! Ao mais novo membro Pitaya, que ainda está por vir, Henrique, a tia já te ama muito e pode ter certeza que farei de tudo para que você cresça em um mundo lindo e cheio de amor.

À minha segunda família, República Pitayas, Dudinha, Gi, Ian, Kakau, Nana e Vanyy (coloquei em ordem alfabética para ninguém brigar), não sei nem por onde começar e o que agradecer, porque foram tantos perrengues, tantas alegrias, tantas vivências ao lado de vocês, que seria impossível resumir em um parágrafo. Mas eu quero agradecer a oportunidade de ser amiga de mulheres admiráveis e transformadoras! Vocês me amaram incondicionalmente e me mostraram que família é a gente que constrói, em nossos moldes, com pessoas que emanam amor e desejam o nosso bem. Então, eu escolhi vocês (na verdade a gente se escolheu) e não poderia ser a melhor escolha de todas. Vocês me ajudaram a me desconstruir e reconstruir; a enxergar minhas capacidades e qualidades; e o mais importante, a ser uma mulher tão extraordinária quanto vocês. Obrigada por serem abrigo e conforto sempre que precisei. É um privilégio poder levar para sempre comigo o riso solto e o companheirismo infindável da Ian; o alto astral e os cuidados afáveis da Gi; as piadas ruins (mas que sempre me faziam rir) e os abraços quentinhos da Vanyy, o carinho e a compreensão ilimitada da Dudinha; os papos aleatórios cheios de gargalhadas e o coração gigantesco da Nana; a playlist que abraça a alma e a bondade gigantesca da Kakau. Vocês me inseriram tão profundamente em suas vidas, que ganhei muitas famílias espalhadas por São Paulo. Obrigada tia Andresa, tio Marcos, Pedro, amiga Fabi, Gabri e dona Romilda por me acolherem com tanto carinho. Amigas, saibam que não importa o tamanho do hiato entre nós, eu sempre estarei aplaudindo vocês de pé, cheia de orgulho, por cada conquista pitayana. Espero reencontrá-las em breve no lugar que vocês merecem estar: No topo do mundo! Às minhas irmãs de alma: AUUUUUUUUUUUUUUUUUUUUU!

A todos que contribuíram de alguma forma, meus mais sinceros: Muito obrigada!

Um beijo minha querida e amada, Unesp!



In nature, nothing exists alone. (CARSON,  
1962)

## RESUMO

A região centromérica dos cromossomos dos eucariotos é composta por sequências repetitivas, sendo representadas principalmente por elementos repetidos em tandem comumente conhecidos como DNAs satélites. Os estudos referentes a essas sequências altamente dinâmicas têm demonstrado seus papéis importantes nos genomas, especialmente no que concerne à evolução genômica, composição cromossômica e diferenciação de espécies. Apesar de sua elevada variabilidade, essas sequências podem apresentar certa conservação entre alguns grupos, o que as tornam atrativas para estudos genômicos, populacionais e evolutivos. Os estudos referentes ao grupo dos peixes que abordam a composição centromérica e a possível conservação de satDNAs nessa região são escassos e ainda pouco explorados em publicações recentes. O presente trabalho envolveu o estudo de um DNA satélite de 177 pb (MelSat03-177) previamente isolado do genoma do peixe Neotropical *Megaleporinus elongatus* (Ordem Characiformes), que está localizado na região centromérica de alguns cromossomos da espécie, no intuito de analisar a possível conservação dessa sequência e conhecer um pouco mais sobre a sua distribuição e organização em outros genomas de peixes, principalmente em Characiformes. Por meio de busca em banco de dados, foram recuperadas sequências homólogas ao satélite MelSat03-177 de outras espécies da Ordem Characiformes, com uma porcentagem relevante de identidade, demonstrando prenúncios de conservação da sequência. O presente trabalho trouxe informações iniciais, mas de grande importância para o estudo de DNAs satélites, especialmente para os satDNA localizados em região centromérica, podendo fomentar diversos estudos relacionados à distribuição, organização e conservação de sequências repetitivas centroméricas presentes no grupo de peixes.

**Palavras-chaves:** Centrômero. SatDNA conservado. Hipótese da biblioteca.

## ABSTRACT

The centromeric region of the eukaryotic chromosomes is composed of repetitive sequences, being represented mainly by tandem repeats commonly known as satellite DNAs. Studies referring to these highly dynamic sequences have demonstrated their important roles in genomes, especially with regard to genomic evolution, chromosomal composition and species differentiation. Despite their high variability, these sequences may show some conservation among some groups, which makes them attractive for genomic, population and evolutionary studies. Studies referring to the group of fish that address the centromeric composition and possible conservation of satDNAs in this region are scarce and still few explored in recent publications. The present work involved the study of a 177 bp satellite DNA (MeISat03-177) previously isolated from the genome of the Neotropical fish *Megaleporinus elongatus* (Order Characiformes) which located in the centromeric region of some chromosomes of the species, in order to analyze the possible conservation of this sequence and to know a few more about its distribution and organization in other fish genomes, mainly in Characiformes. Through a database search, homologous sequences to the satellite MeISat03-177 were retrieved from other species of the Order Characiformes, with a relevant percentage of identity, showing signs of conservation of the sequence. The present work provided initial information, but of great importance for the study of satellites DNA, especially for satDNA in centromeric region, being able to promote several studies related to the distribution, organization and conservation of centromeric repetitive sequences present in the fish group.

**Keywords:** Centromere. SatDNA conserved. Library hypothesis.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	12
1.1	Considerações iniciais sobre sequências repetitivas centroméricas .....	12
1.2	Os DNAs satélites.....	14
1.3	Considerações sobre o organismo de estudo: <i>Megaleporinus elongatus</i> ... .....	16
1.4	Apontamentos sobre o DNA satélite de 177 pb isolado de <i>M. elongatus</i> ..... .....	19
1.5	Objetivos .....	20
<b>2</b>	<b>MATERIAIS E MÉTODOS</b> .....	20
2.1	Materiais.....	20
2.2	Métodos.....	21
2.2.1	<i>Busca em banco de dados</i> .....	21
2.2.2	<i>Análise comparativa via RepeatExplorer e Geneious</i> .....	21
2.2.3	<i>Nomenclatura</i> .....	22
2.2.4	<i>Busca por elementos transponíveis (TEs)</i> .....	23
2.2.5	<i>Busca por motifs</i> .....	23
2.2.6	<i>Alinhamento</i> .....	23
2.2.7	<i>Construção de rede de haplótipos</i> .....	23
2.2.8	<i>Análise dos materiais de localização cromossômica</i> .....	24
<b>3</b>	<b>RESULTADOS</b> .....	24
<b>4</b>	<b>DISCUSSÃO</b> .....	32
4.1	Conservação de uma superfamília de DNA satélite .....	32
4.2	Indícios de conservação da localização cromossômica .....	35
4.3	Função centromérica dos satDNAs estudados .....	37
<b>5</b>	<b>CONCLUSÃO</b> .....	39
	REFERÊNCIAS.....	40

## 1 INTRODUÇÃO

### 1.1 Considerações iniciais sobre sequências repetitivas centroméricas

Os centrômeros correspondem à região constricta dos cromossomos que preservam a ligação entre as duas cromátides-irmãs durante a divisão celular. São tidos como fundamentais para a vida dos eucariotos, uma vez que são responsáveis pela ligação do fuso mitótico e segregação ordenada dos cromossomos homólogos para cada célula-filha durante a divisão celular (ALBERTS, 2017; GARRIDO-RAMOS *et al.*, 1994; METZNER, 1894 *apud* RHOADES, 1940; PLOHL *et al.*, 2008). A deleção ou mutação em suas proteínas associadas podem causar sérios danos às células eucarióticas. Nos humanos, por exemplo, as anormalidades em sua função podem resultar em doenças ou até letalidade (MELTERS *et al.*, 2013; PLOHL; MESTROVIC; MRAVINAC, 2014).

A cromatina centromérica possui uma estrutura distinta da eucromatina e da heterocromatina, uma vez que é constituída pela CENP-A (proteína centromérica-A), uma variante de histona H3 evolutivamente conservada e exclusiva do centrômero. Junto de outras proteínas que compactam os nucleossomos, a CENP-A atua como base para a formação do cinetocoro, um conjunto de complexos multiproteicos necessários para a ligação do fuso mitótico (ALBERTS, 2017; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021).

Proteínas centroméricas como CENP-B, CENP-C, CENP-S, entre outras, são conhecidas por se associarem a sequências de DNA. Em respeito a CENP-B, sabe-se que essa proteína reconhece e se liga em sequências de DNA que possuem uma sequência *motif* de 17 pares de base, denominada caixa CENP-B (FUJITA *et al.*, 2015). A caixa CENP-B foi descrita pela primeira vez em mamíferos - no DNA satélite alfa, porém alguns trabalhos apresentaram sequências semelhantes à caixa CENP-B em outros grupos, como em anfíbios (EDWARDS; MURRAY, 2005), nematóides (MESTROVIC *et al.*, 2013), peixes (RODRÍGUEZ *et al.*, 2017), entre outros (HERRÁN, 2008; DUMONT; FACHINETTI, 2017).

Apesar de conhecido como uma estrutura altamente conservada em eucariotos e com uma função bem definida, o centrômero se insere em um cenário paradoxal no qual apresenta também uma rápida evolução das sequências de DNA que compõem

toda sua região (HENIKOFF; AHMAD; MALIK, 2001; KURSEL; MALIK, 2018, TEIXEIRA *et al.*, 2018). Assim sendo, o centrômero é caracterizado tanto por sua herança estável quanto por sua composição variável. Essa variabilidade se deve principalmente aos elementos repetitivos do tipo DNAs satélites (satDNAs) e elementos transponíveis (TEs), tidos como os principais componentes da heterocromatina centromérica dos eucariotos (PLOHL; MESTROVIC; MRAVINAC, 2014; PLOHL *et al.*, 2008; YU *et al.*, 2017).

Atualmente satDNAs têm sido relacionados a papéis importantes da célula, como por exemplo no funcionamento dos centrômeros (RODRÍGUEZ *et al.*, 2017; RUIZ-RUANO *et al.*, 2016), e são conhecidos por serem um dos componentes mais dinâmicos de um genoma, sofrendo modificações por meio da evolução em concerto (o que explicaria a variação das sequências de DNA na região centromérica) (DOVER, 1982; HENIKOFF; AHMAD; MALIK, 2001).

A funcionalidade dos DNAs satélites vem sendo associada a algumas características de suas sequências, como o tamanho dos monômeros, por exemplo. Seu tamanho pode estar relacionado à organização nucleossômica, à condensação de heterocromatina, à regulação da transcrição e à função centromérica (MELTERS *et al.*, 2013). Posto isso, considerando que o CENH3 (histona específica do centrômero) é essencial para a nucleação do cinetocoro e que determinados elementos repetitivos são flanqueados em sua cromatina (revisado por GARRIDO-RAMOS, 2015; HENIKOFF; AHMAD; MALIK, 2001), alguns estudos sugeriram que o tamanho do monômero de repetição dos satDNAs centroméricos se assemelha ao tamanho de DNA nucleossômico, sendo em torno de 170 pares de base (pb), ou 340 pb, o que corresponde ao tamanho de dois nucleossomos (HENIKOFF; AHMAD; MALIK, 2001; MELTERS *et al.*, 2013; PLOHL; MESTROVIC; MRAVINAC, 2012). Dessa forma, é notável a semelhança encontrada das unidades repetitivas centroméricas entre os mais variados organismos, como por exemplo o alpha-satélite em primatas com 171 pb, no milho com 180 pb e no peixe *Sparus aurata* com 186 pb (HENIKOFF; AHMAD; MALIK, 2001; PLOHL *et al.*, 2008; RODRÍGUEZ *et al.*, 2017).

Além do tamanho semelhante ao DNA nucleossômico, a presença de *motifs* semelhantes a caixa CENP-B em sequências de DNA satélite também podem indicar uma possível função centromérica dessas sequências (LOUZADA *et al.*, 2020; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021).

## 1.2 Os DNAs satélites

As sequências de DNA repetitivo foram apresentadas pela primeira vez por Britten e Kohne (1968), no qual os autores relacionaram diferentes frequências de renaturação do DNA ao grau de repetitividade dessas sequências genômicas. Sendo assim, por meio da renaturação do DNA, as sequências poderiam ser identificadas como DNA altamente repetitivo, moderadamente repetitivo e de cópia única (BRITTEN; KOHNE, 1968; revisado por GARRIDO-RAMOS, 2017; LÓPEZ-FLORES & GARRIDO-RAMOS, 2012). Entre as sequências com DNA altamente repetitivo encontram-se os DNAs satélites. O termo “DNA satélite” (satDNA) surgiu por meio de análises de ultracentrifugação de gradiente de densidade do DNA genômico de um roedor, que demonstraram um componente secundário que poderia se encontrar acima (DNA rico em AT) ou abaixo (DNA rico em GC) da banda correspondente ao volume do genoma (KIT, 1961; SUEOKA, 1961; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021).

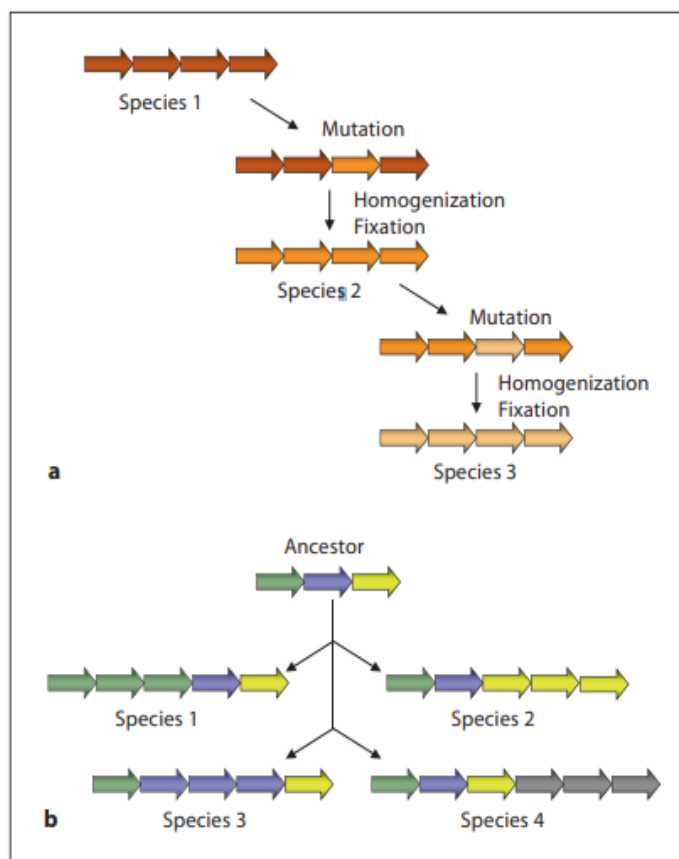
Os satDNAs são sequências organizadas em tandem, e, em sua maioria, não-codificantes, podendo em muitos casos exceder 50% da composição de todo o genoma de um organismo (revisado por GARRIDO-RAMOS, 2017; LÓPEZ-FLORES & GARRIDO-RAMOS, 2012; PLOHL; MESTROVIC; MRAVINAC, 2012). Estão localizados principalmente nas regiões pericentroméricas, centroméricas e subteloméricas dos cromossomos. Possuem um alto grau de polimorfismo, podendo apresentar unidades de repetições (monômeros) de tamanhos variáveis, com dezenas ou até mesmo centenas de nucleotídeos (LÓPEZ-FLORES; GARRIDO-RAMOS, 2012; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021). São organizados dentro das espécies em diferentes famílias, as quais podem ser compartilhadas entre espécies relacionadas ou serem espécie-específicas (revisado por GARRIDO-RAMOS, 2017). A evolução das sequências satélites pode ser explicada por modelos evolutivos como a evolução em concerto e a hipótese da biblioteca.

Evolução em concerto apresenta a concepção na qual componentes de uma família de satélites possuem baixa taxa de divergência, uma vez que seus monômeros são homogeneizados e fixados dentro de uma espécie. Ou seja, apesar de passíveis de mutações nas cópias das sequências satélites de forma independente, a semelhança entre os monômeros é assegurada pela homogeneização (revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021). Esse modelo ocorre por meio do processo

evolutivo de “*molecular drive*”, que corresponde a uma variedade de mecanismos moleculares de *turnover* como: múltiplas inserções de elementos transponíveis, derrapagem de replicação, replicação por círculo rolante, *crossing-over* desigual e conversão gênica (DOVER, 1982; PLOHL; MESTROVIC; MRAVINAC, 2012, revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021; UGARKOVIC; PLOHL, 2002) (Figura 1 a).

A “hipótese da biblioteca” fundamenta-se na ideia de que espécies evolutivamente próximas compartilham uma biblioteca de satDNAs conservada, originadas de um ancestral comum, mas amplificada em cada espécie de forma independente (CAMACHO *et al.*, 2021; FRY; SALSER, 1977; MESTROVIC; PLOHL; MRAVINAC, 1998). Isso faz com que uma sequência contida nessa biblioteca possa aparecer abundante em uma espécie e em poucas cópias nas espécies irmãs (CAMACHO *et al.*, 2021; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021) (Figura 1 b).

**Figura 1** - Representação dos principais modelos evolutivos para as sequências de DNA satélite. Evolução em concerto (a) e “Hipótese da biblioteca” (b).



Fonte: Plohl, Mestrovic e Mravinac (2012).



### 1.3 Considerações sobre o organismo de estudo: *Megaleporinus elongatus*

Peixes neotropicais possuem uma grande diversidade, que pode ser facilmente vista dentro da Ordem Characiformes com uma das famílias mais ricas em espécies, a família Anostomidae. O grupo dos anostomídeos possui uma ampla distribuição em ambientes de água doce da América do Sul tropical e subtropical, sendo constituído por 15 gêneros e apresentando cerca de 150 espécies descritas atualmente (RAMIREZ; BIRINDELLI; GALETTI, 2017; RAMIREZ *et al.*, 2017; REIS; KULLANDER; FERRARIS JR, 2003; SIDLAUSKAS; VARI, 2008; SIDLAUSKAS *et al.*, 2021). São conhecidos popularmente como aracus na região Norte do Brasil e como piaus nas demais regiões (GÉRY 1977 *apud* DE BARROS, 2017). As mais variadas morfologias são encontradas dentro da família, com padrões que vão desde fusiformes e alongados até *deep-bodied* e comprimidos lateralmente (SIDLAUSKAS; VARI, 2008). A respeito de suas características cariotípicas, o grupo apresenta-se conservado em estrutura com número diploide de  $2n = 54$  e morfologia cromossômica semelhante, com cromossomos metacêntricos e submetacêntricos (GALETTI; CESAR; VENERE, 1991), com um único par de cromossomos portadores de RON (região organizadora do nucléolo) para a maioria das espécies (DE BORBA *et al.*, 2013; GALETTI; CESAR; VENERE, 1991).

O gênero *Leporinus* é o mais diverso da família, com aproximadamente 90 espécies descritas. Recentemente o grupo passou por uma revisão e algumas de suas espécies foram reclassificadas em um novo gênero denominado *Megaleporinus* (RAMIREZ; BIRINDELLI; GALETTI, 2017). A nova classificação usou como base a hipótese proposta inicialmente por Galetti Jr, Lima e Venere (1995), que define a existência de um clado monofilético caracterizado pela presença de um sistema de cromossomos sexuais com heterogamia feminina (ZZ/ZW). Características morfológicas e moleculares que destoavam das outras espécies do gênero *Leporinus* também foram utilizadas para separar as linhagens. O novo grupo, *Megaleporinus*, passou a ser constituído por 10 espécies, das quais nove eram classificadas anteriormente no gênero *Leporinus* (*L. obtusidens*, *L. macrocephalus*, *L. elongatus*, *L. muyscorum*, *L. piavussu*, *L. reinhardti*, *L. brinco*, *L. trifasciatus* e *L. conirostris*) e uma no gênero *Hypomasticus* (*H. garmani*) (RAMIREZ; BIRINDELLI; GALETTI, 2017).

Os membros do gênero *Megaleporinus* são maiores do que as espécies do gênero *Leporinus*, o que explica em seu nome o prefixo *mega* (origem grega), que

significa grande ou maior. Apresentam morfologia variada quanto aos tipos de boca, possuem manchas escuras (de 1 a 4) na parte médio-lateral do corpo, três dentes no osso pré-maxilar, três ou quatro dentes no osso dentário e, conforme citado anteriormente, dispõem de um sistema de cromossomos sexuais ZZ/ZW. A maioria se alimenta de plantas, porém inclui também em sua alimentação alguns invertebrados como moluscos, crustáceos e insetos (RAMIREZ; BIRINDELLI; GALETTI, 2017).

Quanto ao cromossomo heteromórfico, sabe-se que a morfologia semelhante entre os cromossomos W das espécies do gênero, quanto ao tamanho, banda C positiva e posição do centrômero, indica que tenham se originado de um ancestral comum (GALETTI; FORESTI, 1986; GALETTI JR; LIMA; VENERE, 1995). Acredita-se que a primeira etapa da diferenciação do cromossomo sexual (W) em *Megaleporinus* seja o processo de heterocromatinização, no qual ocorre o acúmulo substancial de sequências repetitivas (GALETTI; FORESTI, 1986; GALETTI; LIMA; VENERE, 1995; POLTRONIERI *et al.*, 2014).

Diversos dados citogenéticos relacionados à composição do cromossomo sexual W, demonstraram que a distribuição e o conteúdo dessas sequências repetitivas são diferentes entre as espécies do gênero, o que indica uma grande complexidade no desenvolvimento do cromossomo heteromórfico (DE BARROS *et al.*, 2018; CREPALDI; PARISE-MALTEMPI, 2020; DULZ *et al.*, 2020; MOLINA; GALETTI JR, 2007; PARISE-MALTEMPI *et al.*, 2007; POLTRONIERI *et al.*, 2014).

Devido ao acúmulo expressivo de heterocromatina no cromossomo W, o gênero *Megaleporinus* tem sido foco de muitos estudos referentes à identificação, isolamento e caracterização de sequências repetitivas. Parise-Maltempi e colaboradores (2007), por exemplo, isolaram uma sequência repetitiva de 650 pb em *M. elongatus*, chamada LeSpel, pela qual indicaram um possível sistema múltiplo de cromossomos sexuais, devido a presença de marcações semelhantes entre o W e um dos cromossomos de um par de autossomos, sendo esse o par portador da RON. A identificação do elemento repetitivo LeSpel inspirou trabalhos subsequentes relacionados à análises comparativas entre as espécies do gênero *Megaleporinus*. A exemplo disso, Hashimoto e colaboradores (2009) utilizaram LeSpel para identificar híbridos de *M. elongatus* e *M. macrocephalus*, enquanto Marreta, Faldoni e Parise-Maltempi (2012) identificaram padrões de marcação de LeSpel muito semelhantes

entre os cromossomos W ( $W_1$ ) de *M. macrocephalus*, *M. elongatus* e *M. obtusidens*, indicando sua potencialidade como marcador sexo-específico.

Crepaldi e Parise-Maltempi (2020) verificaram que dos 15 satDNAs mapeados em fêmea de *M. elongatus*, 3 se encontravam compartilhados entre autossomos e o sexual ( $W_1$ ), e 9 estavam presentes apenas no  $W_1$ . Utsunomia e colaboradores (2019) apresentaram em seus dados 14 satDNAs hibridados especificamente no cromossomo W de *M. macrocephalus* e 8 localizados tanto no W quanto em alguns autossomos.

A espécie foco do presente trabalho, *Megaleporinus elongatus* (Figura 2), é um peixe Neotropical, endêmico dos rios Jequitinhonha e Pardo, sendo conhecido popularmente como piapara (PERROTTI *et al.*, 2018).

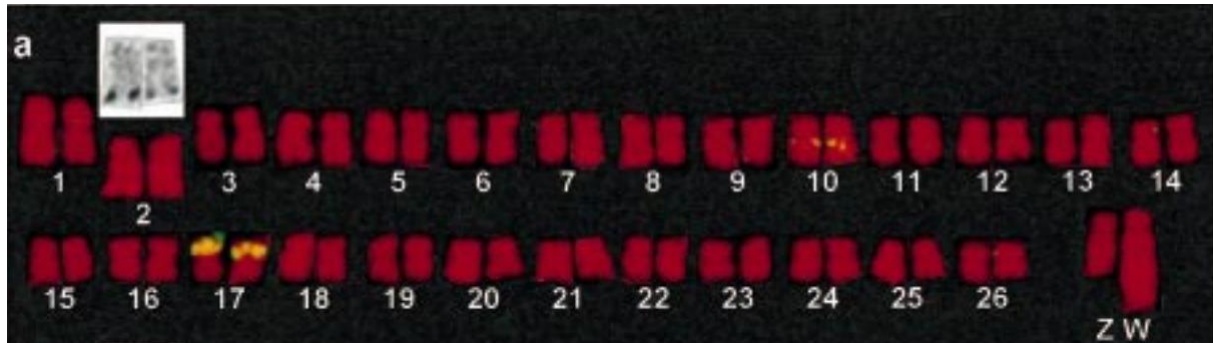
**Figura 2** - Fêmea de *Megaleporinus elongatus*.



Fonte: Adaptado de Crepaldi (2019).

Possui número diploide  $2n = 54$ , conforme esperado para as espécies da família (MARTINS; GALETTI, 1999), e um hipotético sistema sexual múltiplo do tipo  $Z_1Z_1Z_2Z_2/W_1Z_1W_2Z_2$  (PARISE-MALTEMPI *et al.*, 2007), no qual a fêmea apresenta um cromossomo heteromórfico ( $W_1$ ), subteloentríco e altamente heterocromático e diferenciado (GALETTI; FORESTI, 1986) (Figura 3). A taxa elevada de heterocromatina encontrada no braço maior desse cromossomo condiz com a grande quantidade de sequências de DNAs repetitivos, como DNAs satélites, presentes nessa região (CREPALDI; PARISE-MALTEMPI, 2020).

**Figura 3** - Cariótipo correspondente à fêmea de *Megaleporinus elongatus*, demonstrando o número diplóide encontrado para a espécie  $2n=54$ . Ao lado superior esquerdo, evidencia-se a RON presente no par 2 (caixa branca); marcação do 5S rDNA nos pares 10 e 17; e par sexual ZW.



Fonte: Adaptado de Martins e Galetti (1999).

#### 1.4 Apontamentos sobre o DNA satélite de 177 pb isolado de *Megaleporinus elongatus*

O terceiro satélite mais abundante do genoma da fêmea de *M. elongatus*, MeISat03-177 (anteriormente denominado MeISat07-177 por Crepaldi e Parise-Maltempi, 2020 e atualizado recentemente por Crepaldi *et al.*, 2021), está clusterizado na região centromérica de 4 pares autossômicos e do cromossomo heteromórfico ( $W_1$ ). Este elemento repetitivo traz uma característica relevante em relação ao seu tamanho de 177 pb, no qual se assemelha ao tamanho normalmente encontrado para os monômeros repetitivos centroméricos já publicados para outros organismos (EDWARDS; MURRAY, 2005; GARRIDO-RAMOS *et al.*, 1994; HERRÁN *et al.*, 2001; HERRÁN *et al.*, 2008; RODRÍGUEZ *et al.*, 2017).

Essa mesma sequência mostrou semelhança com outros satDNAs descritos para peixes, como por exemplo o satélite MmaSat047-177, encontrado na espécie congênere *M. macrocephalus* (CREPALDI; PARISE-MALTEMPI, 2020), e CgoSat44-177 em *Characidium gomesi* (dados não-publicados), ambos também possuindo 177 pb (SERRANO-FREITAS *et al.*, 2020; UTSUNOMIA *et al.*, 2019). Assim como demonstrado para MeISat03-177 (CREPALDI; PARISE-MALTEMPI, 2020), o trabalho de Serrano-Freitas e colaboradores (2020) apresentou marcações do satélite CgoSat44-177 em regiões centroméricas de apenas alguns pares cromossômicos.

De forma interessante, Crepaldi e Parise-Maltempi (2020) demonstraram uma distribuição diferente do satélite MelSat03-177 nas regiões centroméricas entre as espécies. Na espécie *M. elongatus* os sinais de FISH se mostraram em 9 cromossomos e em *Leporinus friderici*, uma espécie evolutivamente próxima, apareceram em apenas 2 cromossomos (CREPALDI; PARISE-MALTEMPI, 2020). Esse mesmo cenário foi observado entre sete espécies de Characidae estudadas por Utsunomia *et al.* (2017), que apresentaram marcações distintas nas regiões centroméricas para o satélite MsaSat01-177. O satélite apresentou-se acumulado, por exemplo, em 1 par cromossômico de *Astyanax bockmanni* e em 18 pares de *Moenkhausia sanctaefilomenae* (UTSUNOMIA *et al.*, 2017). Esse aspecto traz reflexões interessantes sobre esses satélites, pois ao mesmo tempo que essas sequências são consideradas conservadas, percebe-se também uma diversificação de sua distribuição entre espécies correlatas.

Diante das características apresentadas quanto à conservação e distribuição, o satélite MelSat03-177 e os satélites relacionados tornaram-se componentes interessantes a serem estudados, pois além de mostrar aspectos importantes relacionados aos DNA satélites, podem trazer informações relevantes sobre as regiões centroméricas dos cromossomos de peixes e de vertebrados em geral.

## **1.5 Objetivos**

O projeto tem como objetivo conhecer melhor o satélite MelSat03-177, isolado do genoma de fêmea de *Megaleporinus elongatus* por Crepaldi e Parise-Maltempi (2020). Assim, tem como meta avaliar indícios do satDNA em outros representantes da Ordem Characiformes, utilizando ferramentas de bioinformática; levantamento de dados disponíveis em bancos de dados públicos; e materiais de localizações cromossômicas previamente publicados. Esses dados serão analisados, com o intuito de verificar o caráter conservado de uma sequência repetitiva de região centromérica em peixes.

## **2 MATERIAIS E MÉTODOS**

### **2.1 Materiais**

Para a realização das análises foi utilizada como base a sequência de DNA satélite previamente isolada, MelSat07-177 (com código de acesso MK977865.1), a qual foi renomeada recentemente como MelSat03-177 por Crepaldi *et al.* (2021). Sendo assim, no presente trabalho, foi utilizada a nomenclatura atualizada do satDNA.

O satélite em estudo foi obtido por Crepaldi e Parise-Maltempi (2020) a partir das análises do sequenciamento de baixa cobertura do genoma da fêmea de *Megaleporinus elongatus*, na plataforma Illumina (Illumina Inc., San Diego, CA, USA).

A validação da repetição em tandem foi realizada por meio dos *softwares RepeatExplorer* (NOVÁK *et al.*, 2013) e *Geneious v4.8.*; e a sequência de seu monômero foi definida pelo programa *Mega v8.0*, sendo ela composta por 177 pb: 'CAATTCAGAACCAAATTGACACTTTTCACAAGTTTGTTCATCGCATTACTCCAAAT TCCTACAGTAAAATTACTTGTTTCATCATAGCATGGGTACAAGAGAGAGTGTTCTT CAGGAATATGCTGGTATAATTTGATAACAATTTGATTTGTCTATTTTATGGGACCT CAAATTGTGAT' (CREPALDI; PARISE-MALTEMPI, 2020).

## 2.2 Métodos

O desenho experimental desenvolvido para o trabalho foi dividido em 8 etapas: (1) busca em banco de dados; (2) análise comparativa via *RepeatExplorer* e *Geneious*; (3) nomenclatura; (4) busca por elementos transponíveis (TE); (5) busca por *motifs* de CENP-B; (6) alinhamento entre as sequências encontradas; (7) construção de rede de haplótipos; (8) análise de dados prévios quanto à localização cromossômica.

### 2.2.1 Busca em banco de dados

Usando os parâmetros *default* dos programas BLAST (*Basic local alignment search tool*) (<http://www.ncbi.nlm.gov/Blast/>) e Ensembl (<https://www.ensembl.org/index.html>) foram buscadas similaridades com sequências já depositadas para a sequência MelSat03-177.

Em relação a análise por Ensembl, foram realizadas buscas nos genomas de 87 espécies de peixes ósseos, das quais 2 pertenciam à ordem Characiformes.

### 2.2.2 Análise comparativa via *RepeatExplorer* e *Geneious*

Para encontrar sequências homólogas ao MelSat03-177 em *Astyanax fasciatus*, *Astyanax paranae* e *Moenkhausia sanctaefilomenae*, os genomas dessas espécies foram primeiramente recuperados da plataforma *Sequence Read Archive* (SRA) (códigos de acesso: SRR8476332, SRR5461470, SRR8476336, respectivamente) e, posteriormente, inseridos no *RepeatExplorer* (na plataforma Galaxy: <https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>) para a realização de uma rodada de clusterização.

Os arquivos recuperados do SRA foram inseridos no “*Preprocessing of FASTQ paired-end reads*” e como resultado foram obtidos os documentos “*Interlaced paired reads*”, para cada um dos genomas.

As sequências recuperadas por BLAST, por Ensembl e a de MelSat03-177 foram adicionadas como uma biblioteca extra para ocorrer concomitantemente ao processo de clusterização, a fim de conferir quais clusters resultantes poderiam ter alguma porcentagem de identidade com as sequências adicionadas. Para isso, por meio do *RepeatExplorer2 clustering* foram adicionados os arquivos “*Interlaced paired reads*”, no campo *NGS reads* e as sequências supracitadas (no formato FASTA) no campo “*Custom repeat database*”.

Os *clusters* que apresentaram, no *output* do *RepeatExplorer*, resultados pertinentes de identidade foram analisados no *software* Geneious v4.8.5 a fim de validar as repetições em tandem. Então, foram realizadas análises de gráficos *Dotplot* de cada um dos *contigs* contidos em cada *cluster* para seleção daqueles com aspecto de repetição em tandem.

As sequências dos *contigs* selecionados foram inseridas na plataforma *Tandem Repeats Finder* (<https://tandem.bu.edu/trf/trf.html>) para construção da sequência consenso de cada um deles. Para o caso em que o *cluster* apresentava duas consensos, foi escolhida a sequência correspondente ao *contig* maior.

### 2.2.3 Nomenclatura

A nomenclatura dos satélites isolados foi inspirada na que foi proposta por Ruiz-Ruano e colaboradores (2016), porém não a seguiu com exatidão, devido a ausência de dados. Como não foram estudados todos os potenciais satélites de cada espécie, não foi possível obter os dados numéricos em ordem crescente, correspondentes à abundância dos satélites. Sendo assim, a nomenclatura dos satélites, utilizada no

referido trabalho, apresentou os códigos de abreviação de nome da espécie, seguido do tamanho da unidade do repetitivo.

#### 2.2.4 Busca por similaridade com elementos transponíveis (TEs)

As sequências obtidas nos bancos de dados (BLAST e Ensembl), as isoladas no *RepeatExplorer* e a de MelSat03-177 foram buscadas via *Replibase*, utilizando a ferramenta CENSOR (<https://www.girinst.org/replibase/>), a fim de encontrar alguma relação ou associação com outros repetitivos - elementos transponíveis (TEs).

#### 2.2.5 Busca por motivos

Buscou-se em todas as sequências recuperadas de satDNA o *motif* de ligação de 17 pb de CENP-B (NTTCGNNNNANNCGGGN), por meio da plataforma The MEME Suite, utilizando a ferramenta FIMO (versão 5.4.1) (<https://meme-suite.org/meme/tools/fimo>).

#### 2.2.6 Alinhamento

As sequências de DNA satélite recuperadas foram alinhadas por meio da ferramenta MUSCLE (EDGAR, 2004) do *software* Geneious v4.8.5, com parâmetros padrão, para a verificação dos graus de similaridade entre elas. Dessa forma, cada sequência recuperada foi alinhada primeiramente com o MelSat03-177 e, posteriormente, foram alinhadas em conjunto.

Por meio do alinhamento também foi gerado um logotipo de sequência, utilizando o mesmo programa, para a análise de conservação da sequência de satDNA entre as espécies.

#### 2.2.7 Construção de rede de haplótipos

Para a construção da rede de haplótipos foi utilizado o alinhamento de todas as sequências satélites em conjunto como *input* do programa PHYLOViZ 2.0 (NASCIMENTO *et al.*, 2017), em sua versão online (<http://online.phyloviz.net/index>).



Além disso, por meio do *software* Geneious v4.8.5 cada satDNA foi etiquetado com o nome de sua respectiva espécie e família.

Após anotação, foi exportado um arquivo em formato FASTA contendo as sequências de DNA satélite e as informações inseridas para cada uma, em forma de tabela. O documento exportado também foi inserido no PHYLOViZ, sendo usado como dado auxiliar para a construção da rede.

### 2.2.8 Análise dos materiais de localização cromossômica

Para os satDNAs analisados foram buscados na literatura dados que demonstrassem a localização cromossômica dessas sequências em específico.

## 3 RESULTADOS

O resultado obtido nas buscas realizadas através do banco de dados NCBI BLAST demonstrou que o satélite MelSat03-177 apresentou semelhanças com os satélites isolados na espécie *Megaleporinus macrocephalus*, MmaSat047-177, com *percent identity* de 92.97% e *e-value*  $3e-44$  (com código de acesso MG819041.1), e *Characidium gomesi*, CgoSat44-177, com *percent identity* de 77.44% e *e-value*  $1e-18$  (com código de acesso MG764457.1). A sequência de CgoSat44-177 possui 177 bp, confirmada pelo trabalho em que foi isolada (SERRANO-FREITAS *et al.*, 2020) e se encontra duplicada no banco de dados com 354 pb, o que explica o seu uso para as análises de alinhamento e homologia.

Em relação às consultas realizadas no Ensembl, das 87 espécies de peixes analisadas, 3 apresentaram resultados relevantes de identidade com a sequência MelSat03-177, sendo elas *Pygocentrus nattereri*, *Astyanax mexicanus* e *Electrophorus electricus* (com código de acesso MAUM01316306:9-182, APWO02000820:312244-312310 e RBHW02000401:130304-130430, respectivamente), dado que as duas primeiras pertencem também à Ordem Characiformes e a última à Ordem Gymnotiformes. Ainda seguindo a ordem, as sequências apresentaram valores de porcentagem de identidade 82.18%, 85.19% e 81.25%, e *e-value*  $3e-19$ ,  $5e-05$  e  $9e-08$ .

Da iteração realizada pelo *RepeatExplorer* para os três genomas inseridos foram apresentados *clusters* com porcentagem de identidade para as sequências de

MelSat03-177 e as recuperadas acima utilizadas como bibliotecas extras. O *output* do genoma de *Astyanax fasciatus*, demonstrou três *clusters* com resultados significativos, assim o *cluster* 32 (CL 32) demonstrou similaridade com a sequência de satDNA de *P. nattereri* (0.16%), o *cluster* 72 (CL 72) com *P. nattereri* (65.91%), *E. electricus* (4.87%), *M. elongatus* (1.95%), *A. mexicanus* (1.30%), *C. gomesi* (0.32%) e *M. macrocephalus* (0.32%), e o *cluster* 237 (CL 237) com *E. electricus* (20.75%) e *P. nattereri* (1.89%). O *output* de *Astyanax paranae* apresentou para o *cluster* 185 (CL 185) similaridade com os satDNA de *P. nattereri* (32.35%), *A. mexicanus* (26.47%) e *M. macrocephalus* (2.94%); e o de *Moenkhausia sanctaefilomenae* apresentou para o *cluster* 146 (CL 146) similaridade com os satDNA de *A. mexicanus* (25.51%), *P. nattereri* (11.22%) e *M. macrocephalus* (8.16%).

Os cinco *clusters* (CL 32, CL 72, CL 237, CL 185 e CL146) foram analisados para a validação da repetição em tandem. Como resultado, apenas CL 72 e CL 146, correspondentes às espécies *A. fasciatus* e *M. sanctaefilomenae*, respectivamente, apresentaram características de DNA satélite. Com isso, foram construídas as sequências consensos apenas para estes *clusters*. No caso de *A. fasciatus*, foram obtidas sequências de duas variantes, no entanto utilizamos para o estudo apenas a sequência correspondente ao *contig* maior.

Em suma, por meio de buscas em bancos de dados e análises comparativas no *RepeatExplorer* foram recuperadas um total de sete sequências similares ao satDNA MelSat03-177, visto que as sequências recuperadas pelo banco de dados Ensembl foram consideradas como satDNAs putativos. Além disso, seguindo o objetivo do trabalho, todas as espécies selecionadas pertencem à Ordem Characiformes, com exceção da espécie *Electrophorus electricus*. No entanto, a sequência recuperada de *E. electricus* foi inserida no estudo por demonstrar dados relevantes quanto à identidade. O nome e a sequência consenso de cada sequência de satDNA isolada e recuperada estão representados no Quadro 1.

**Quadro 1** - DNAs satélites recuperados dos bancos de dados (BLAST e Ensembl) e isoladas por *RepeatExplorer*, com suas respectivas sequências consenso.

SatDNA	Sequência consenso
MmaSat047-177	ATCACAATTTGAGGTCCCATAAAATAGACAAATCAAGTTGT TGGCAAACATACTAGCATAATCCAGAAAAACACTCTCATT TGTACCCATGTTATGATGAACAAGTAATTTTACTGTAGGAA TTTGGAGTAATGGGATGACAACTTGTGAAAAGTGCAATT CGGTTCTGAATTG
CgoSat44-177	TGTAATAGTGATATGAGCTTCATGTATTTTTACTGTAGGAA TTTGGAGTAATTTGATGGCAAAGTTCAGTAAAATGTCAATT TGGTTATTTATATACCTGTATTTAAAGGACAATAAAATTA ATAATTGATTGTTATCAAACACTACACCTATATATTCTTGTAGT GCACTGTCTCCTGTAATAGTGATATGAGCTTCATGTATTTT TACTGTAGGAATTTGGAGTAATTTGATGGCAAAGTTCAGTA AAATGTCAATTTGGTTATTTATATACCTGTATTTAAAGGACA ATAAAATTAATAAATTGATTGTTATCAAACACTACACCTATAT ATTCTTGTAGTGCACTGTCTCC
EelSat127	TTTGGCATCCCATTACCACCAAATTCCTACAGTAAAAACAC TTGAAACTCATACCCTGGGTACATGAGAGAGTTTTAAAGA ATTTAGTGGAGTGATCTGATAACAATTTGATTTGTTTATTTT AT
PnaSat174	GATAACAATTTGATTTGCTTATTTTATAGGACCTCAAATTGT GATTAATTCATAACCGAATAGACACTTTTCACAAGTTTGCC ATCCATTATTCCAAATTTCTACTGTAATATCACTTGGTCC ACATACCATGAATACAAGAAAGAGTGTTCTTTAAGTATATA CTTGTCTGA
AmeSat67	AACCAAATAGGCATTTTTCCCAAGTTTGGTCTCCCATTACT CCAAAATGCTACAGTAAAATCACTTG
AfaSat177	CTATTCGGTTGTGCATAAACCAGATTTTCGGAGTCCTATAAA ACCAGCAAATCAAATTGTTATCCAATGCCACCAAGACAGT CTTAAAGAACACATTCCATACTAATAATGACATGAGTAACA AAGAATATTACTGTATCATTTTGGAGTAATTGGATACCAA CTTAGGAAAAATGT
MsaSat177	CGAATAGACATTTTTCTAAGTTTGGAAATCCCATTACTCCA AAATGATACAGTAAAATTTACTTGATACACATGACATTAATA CTATGAAATGTGTTCTTCAAGGATTTTCATAGTGGCATTGTA TAACAATTTGATTTTCAGGTTTTATAAGACTCCAAAATCTGT TTTATGCAATAC

Fonte: Dados da pesquisa.

Por meio das análises comparativas com satDNA de *M. elongatus* foi observado que, das 7 sequências recuperadas, três possuem o monômero de mesmo tamanho do satélite MelSat03-177; uma aproxima-se muito com apenas três pares de bases de diferença; duas sequências apresentam tamanho aproximado com diferença entre 40 e 50 pares de bases; e somente uma (AmeSat67) apresenta um tamanho menor do esperado, apresentando menos da metade de 177 pares de bases. Quanto ao conteúdo de A + T dos satDNAs, calculado via Geneious, os valores variam de 62.7% a 70.6% (Tabela 1).

Os alinhamentos entre os consensos (Figura 4 A) dos satDNAs recuperados com MelSat03-177 demonstraram homologias significativas, apresentando dados de *pairwise identity* acima de 70%, exceto para CgoSat44-177, que apresenta 56.6%, como pode ser visto na Tabela 1.

O logotipo de sequência realizado por meio do alinhamento geral, que exibiu 74.1% de identidade entre todas as sequências, apresentou dados interessantes quanto a conservação das sequências apresentadas (Figura 4 B). Dado que, quanto maior for o tamanho da base, mais conservada ela é na sequência.

**Tabela 1** - Dados da análise comparativa das sequências satélites recuperadas com a sequência satélite MelSat03-177, isolada de *Megaleporinus elongatus*. A tabela segue a ordem decrescente do valor de *pairwise identity* (%) em relação à MelSat03-177. Também estão incluídos os valores do número diplóide (2n) das espécies; quantidade de bases dos satDNAs; e porcentagem do conteúdo de A+T.

SatDNA	Espécie	2n	Pares de base (pb)	A+T (%)	<i>Pairwise identity</i> (%)
MelSat03-177 (1)	<i>Megaleporinus elongatus</i>	54	177	66.7	-
MmaSat047-177 (2)	<i>Megaleporinus macrocephalus</i>	54	177	66.1	92.7
AmeSat67	<i>Astyanax mexicanus</i>	50 + B	67	62.7	84.2
PnaSat174	<i>Pygocentrus nattereri</i>	60	174	69.0	82.5
EelSat127	<i>Electrophorus electricus</i>	52 + B	127	66.9	81.3
CgoSat44-177 (3)	<i>Characidium gomesi</i>	50 + B	354	73.3	56.6
MsaSat177	<i>Moenkhausia sanctaefilomenae</i>	50 + B	177	70.6	70.6
AfaSat177	<i>Astyanax fasciatus</i>	46 + B	177	66.7	70.1

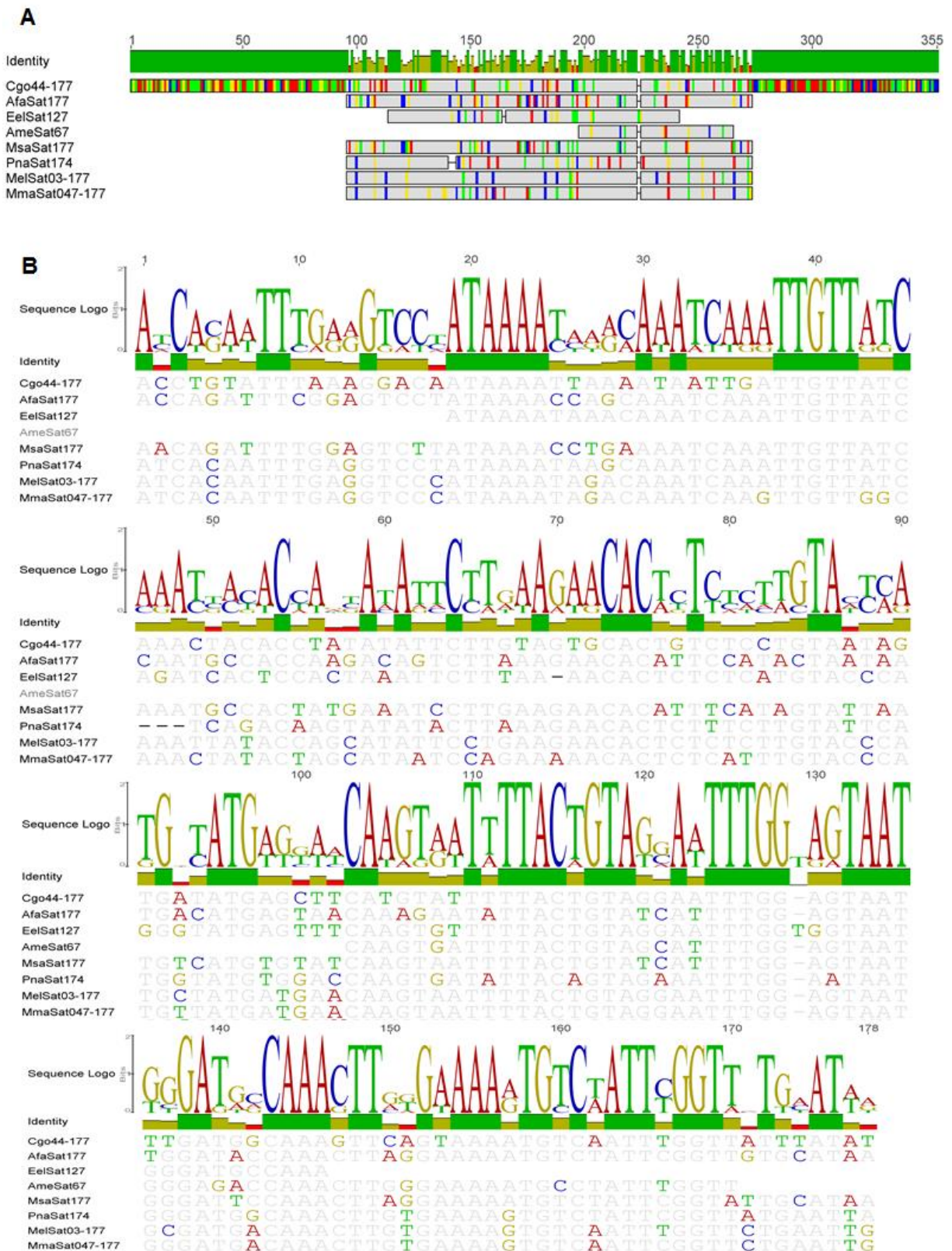
Fonte: Dados da pesquisa.

(1) Crepaldi; Parise-Maltempo (2020)

(2) Utsunomia *et al.* (2019)

(3) Serrano-Freitas *et al.* (2020)

**Figura 4** - Alinhamento das seqüências consensos dos satDNAs estudados. (A) Alinhamento com todas as seqüências, incluindo Cgo44-177 completa. (B) Logotipo de seqüência, seguido de seqüências consensos correspondentes aos satDNAs isolados; dado que, para a seqüência de Cgo44-177 está apresentado apenas o fragmento alinhado com as outras seqüências.

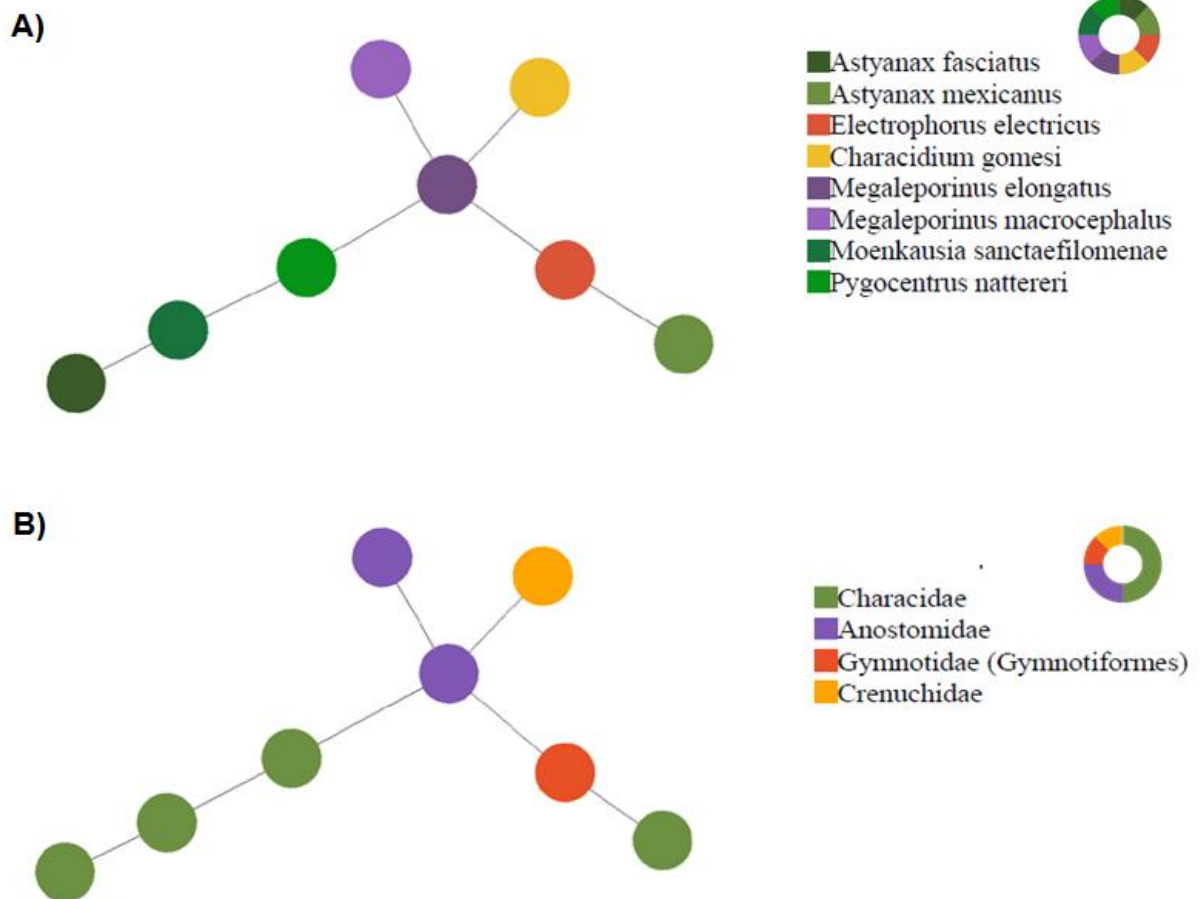


Fonte: Dados da pesquisa

A busca realizada na plataforma *Repbase* apresentou similaridade para os satDNAs AfaSat177 e MsaSat17 com o transposon (DNA8-20\_DR) de *Danio rerio*, e com o retrotransposon LTR (Copia2-I\_Dmoj) da espécie *Drosophila mojavensis*, respectivamente. Ambos os resultados, porém, não apresentaram valores significativos de identidade para maiores inferências quanto a elementos transponíveis associados. Quanto à procura por *motif* de CENP-B, não houve resultados de similaridade para nenhuma das sequências estudadas.

Por meio de um alinhamento geral foi construída uma rede de haplótipos para os satDNAs estudados, também conhecida como árvore geradora mínima (MST - *minimum spanning tree*), em dois cenários diferentes: um representando a relação entre as espécies (Figura 5 A) e outro a relação entre as famílias (Figura 5 B), com o intuito de avaliar a topologia da rede de haplótipos e analisar como as sequências estão se comportando entre si, principalmente no que concerne à famílias diferentes. À exemplo disso, o haplótipo de *Astyanax mexicanus* apresentou uma posição interessante ao localizar-se distante dos outros haplótipos de espécies também pertencentes à família Characidae (Figura 5 A e B). Em relação à família Anostomidae, os haplótipos correspondentes a ela encontram-se próximos, corroborando com o alto valor de identidade entre as sequências de *M. elongatus* e *M. macrocephalus*. O haplótipo de Gymnotidae (Ordem Gymnotiformes) apresenta uma ligação direta ao haplótipo de *M. elongatus*, se comportando de forma diferente do esperado, uma vez que em um primeiro momento imaginávamos que sua sequência se encontraria mais distante por não pertencer à Ordem Characiformes.

**Figura 5** - Rede de haplótipos mostrando a relação entre as sequências estudadas, demonstrando os haplótipos correspondentes às espécies, em A, e às famílias, em B. Note o haplótipo de *A. mexicanus* isolado do restante dos Characidae, e a sequência de *M. elongatus* demonstrando certo grau de homologia com as diversas famílias.



Fonte: Dados da pesquisa.

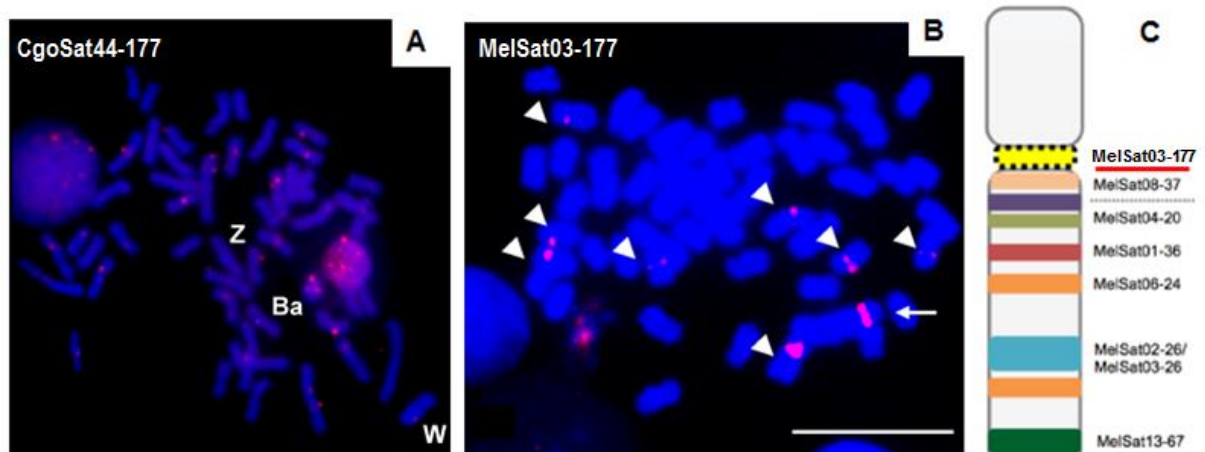
Foram encontrados apenas dois trabalhos envolvendo a localização cromossômica, sendo correspondentes aos satélites CgoSat44-177 e MeISat03-177, os quais foram mapeados por meio de hibridação *in situ* fluorescente (FISH) em preparações cromossômicas de fêmeas de *C. gomesi* e *M. elongatus*, respectivamente (CREPALDI; PARISE-MALTEMPI, 2020; SERRANO-FREITAS *et al.*, 2020).

O satélite CgoSat44-177 apresentou marcações de FISH em aproximadamente metade dos pares autossomos e no cromossomo B, não havendo marcação nos cromossomos sexuais (SERRANO-FREITAS *et al.*, 2020) (Figura 6 A). O satDNA MeISat03-177 apresentou marcações visíveis em 4 pares de cromossomos



autossomos e no cromossomo sexual  $W_1$  (CREPALDI; PARISE-MALTEMPI, 2020) (Figura 6 B). Ambos os satélites apresentaram marcações nas regiões centroméricas dos cromossomos, o que pode ser visto em maior detalhe para o satDNA MeISat03-177 na Figura 6 C.

**Figura 6** - Localizações cromossômicas dos satDNAs. (A) FISH em metáfases mitóticas dos satDNAs isolados em fêmea de *Characidium gomesi*, retirado de Serrano Freitas *et al.* (2020). (B) FISH em metáfases mitóticas dos satDNAs isolados em fêmea de *Megaleporinus elongatus*, retirado de Crepaldi e Parise-Maltempi (2020). (C) Parte do ideograma do cromossomo sexual  $W_1$  da fêmea de *Megaleporinus elongatus*, o qual evidencia a posição centromérica (pontilhado com fundo amarelo) do satDNA MeISat03-177 (sublinhado em vermelho), adaptado de Crepaldi (2019). Dado que, em A, Z = cromossomo sexual Z; W = cromossomo sexual W; Ba = cromossomo B; e em B, as pontas das setas indicam os cromossomos autossomos com sinais de FISH e a seta destaca o cromossomo  $W_1$  com seu respectivo sinal de FISH.



## 4 DISCUSSÃO

### 4.1 Conservação de uma superfamília de DNA satélite

No presente trabalho foram realizadas buscas por sequências homólogas ao satDNA MeISat03-177 de *M. elongatus* em bancos de dados e em genomas de peixes Characiformes já sequenciados, por meio de clusterização, e os resultados obtidos mostram que as sequências recuperadas compõem uma superfamília de DNA satélite,

sendo corroborada pelos valores de identidade (RUIZ-RUANO *et al.*, 2016, 2017). As sequências MmaSat047-177, AmeSat67, PnaSat174 e EelSat127 (referentes às espécies *M. macrocephalus*, *A. mexicanus*, *P. nattereri* e *E. electricus*, respectivamente) indicaram pertencerem à mesma família do satDNA MelSat03-177, com valores de identidade maiores e não foram encontradas sequências que possivelmente constituem-se à mesma variante de MelSat03-177, sendo a sequência de *M. macrocephalus* com maior valor de identidade.

O logotipo de sequência corrobora com o valor de identidade observada no alinhamento geral, para o qual foi demonstrada a presença de pequenos blocos homólogos espalhados entre todas as sequências alinhadas; predominantemente algumas bases em relação a outras em posições específicas; e uma notável similaridade entre os nucleotídeos localizados na porção final das sequências. Por conta disso, denotamos indícios de certo grau de conservação entre as sequências estudadas. Essa variação entre as sequências conservadas também foi observada em outros trabalhos com peixes (DOS SANTOS *et al.*, 2021; UTSUNOMIA *et al.*, 2017), porém em um grau menor de conservação do que foi constatado em nossos resultados.

Como esperado, foram encontradas algumas sequências com tamanho de 177 pb, demonstrando o tamanho comumente observado em unidades repetitivas centroméricas de outros organismos (HENIKOFF; AHMAD; MALIK, 2001; PLOHL *et al.*, 2008; RODRÍGUEZ *et al.*, 2017). Porém, curiosamente, também foram observadas duas sequências com tamanho de monômeros menores, EelSat127 e AmeSat67, que exibiram uma alta homologia com MelSat03-177. Essa divergência observada entre os tamanhos das unidades de repetição é comumente encontrada entre as famílias de satDNA pertencentes a uma superfamília (RUIZ-RUANO *et al.*, 2016, 2018). Além disso, durante a procura por sequências homólogas a MelSat03-177 encontramos na literatura um satDNA centromérico isolado da espécie de peixe *Solea senegalensis*, também com localização e tamanho de monômero (177 pb) esperados (RODRÍGUEZ *et al.*, 2017); no entanto, ao contrário do que se esperava, não foi encontrada homologia significativa dessa sequência com MelSat03-177, mas devido ao seu tamanho e localização é um elemento interessante de ser estudado com maior profundidade em outros genomas.

Quanto à análise filogenética, analisando a topologia da rede de haplótipos gerada, foi observado que o haplótipo de *M. elongatus* está inserido no centro da rede

e é o único que apresenta ligações com haplótipos pertencentes às quatro famílias estudadas. Isso mostrou uma relação mais próxima dessas sequências (MmaSat047-177, CgoSat44-177, PnaSat174 e EelSat127) com o satDNA de *M. elongatus*, além de validar a alta similaridade encontrada entre o monômero de MelSat03-177 com a maior parte das sequências recuperadas (principalmente com o haplótipo de *M. macrocephalus*, pertencente à mesma família, Anostomidae).

A família Characidae apresenta os haplótipos mais próximos, sendo a única com mais passos de mutação em relação a sequência de *M. elongatus* e com maior grau de diferenciação entre as espécies. Isso pode ser interpretado por meio da hipótese da biblioteca, que propõe o compartilhamento de uma biblioteca de satDNAs entre espécies evolutivamente próximas, porém amplificada independentemente dentro de cada espécie (CAMACHO *et al.*, 2021; FRY; SALSER, 1977; MESTROVIC; PLOHL; MRAVINAC, 1998; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021), o que explica os passos de mutação encontrados entre elas. Também pode ser explicado por meio da evolução em concerto, que fundamenta-se na ideia de que as sequências de uma mesma família de satDNAs são semelhantes entre elas (ELDER JR; TURNER, 1995; revisado por GARRIDO-RAMOS, 2017; LORITE *et al.*, 2017; PALACIOS-GIMENEZ *et al.*, 2020; revisado por THAKUR; PACKIARAJ; HENIKOFF, 2021), corroborando com os haplótipos mais próximos e relacionados, observados para a família Characidae.

No entanto, o haplótipo de *A. mexicanus* não se encontra intimamente ligado ao seu congênere *A. fasciatus*, como também não está agrupado com os outros haplótipos da família Characidae. Esse distanciamento de *A. fasciatus*, bem como a sua maior relação com o haplótipo da espécie de uma outra Ordem (Gymnotiformes), levantou questionamentos quanto à evolução destas sequências, uma vez que entendendo a evolução independente aceita para os satDNAs, a localização do haplótipo não reflete necessariamente a relação filogenética entre as espécies (DOS SANTOS *et al.*, 2021). Porém, segundo os modelos evolutivos citados no trabalho, há uma maior relação e semelhança entre as sequências de DNA satélite de espécies próximas, esperando, assim, que os haplótipos correspondentes às espécies relacionadas se encontrem mais próximos, o que não foi observado para AmeSat67. Também foi observado no alinhamento geral que AmeSat67 se alinhou com a porção final das outras sequências, onde os nucleotídeos se apresentam mais conservados, podendo-se inferir que o posicionamento do haplótipo correspondente a AmeSat67 na

rede seja tanto um reflexo do alinhamento, que usou como base a sequência de MeiSat03-177, quanto um possível caso de convergência com a sequência de EelSat127. Contudo, seria necessário um maior aprofundamento nas análises para entender o comportamento destas sequências e validar as nossas hipóteses.

Algumas análises interessantes que poderiam ser aplicadas futuramente na tentativa de explicar a localização de AmeSat67 na rede de haplótipos foram expostas por Camacho *et al.* (2021). Posto que as espécies de gafanhotos *Oedaleus decorus* e *Locusta migratoria* não apresentaram seus haplótipos adequadamente separados, algumas famílias de satDNA de *O. decorus* localizaram-se espalhadas e mais relacionadas com as famílias de satDNA da espécie *L. migratoria*. Esse padrão foi denominado como classificação incompleta da biblioteca (*Incomplete Library Sorting* - ILibS) (CAMACHO *et al.* 2021), entretanto, o estudo com ILibS necessita de mais avanços e outras análises filogenéticas, a fim de conseguir visualizar esse fenômeno em outros contextos de evolução de satDNAs como no caso das sequências estudadas no presente trabalho.

A conservação de satDNAs já foi observada dentro da Ordem Characiformes, com uma família de satDNA pericentroméricos com 177 pb e específica da família Characidae (UTSUNOMIA *et al.*, 2017). Apesar disso, a conservação de satDNA em toda uma ordem não é comumente encontrada, sendo um satDNA de 52 pb, CharSat01-52 pb, o único encontrado entre 200 famílias de satDNA estudadas que está compartilhado em espécies distantemente relacionadas e distribuídas em três famílias da Ordem Characiformes (DOS SANTOS *et al.*, 2021; UTSUNOMIA *et al.*, 2019). Para compreender melhor o indício de conservação encontrado, estudos de divergência e abundância são importantes para entender a evolução e conservação dessas sequências na Ordem Characiformes.

Nossos resultados trazem indícios de um caso incomum de conservação de uma superfamília de satDNA, contudo, como nossas análises são preliminares, precisamos de buscas mais refinadas, a fim de verificar a conservação dessas sequências dentro dos genomas das espécies, uma vez que para a maioria delas só foram realizadas buscas em banco de dados.

## **4.2 Indícios de conservação da localização cromossômica**

Os dados sobre a localização cromossômica encontrados para os dois satDNAs com mapeamento cromossômico, CgoSat44-177 e MelSat03-177 (CREPALDI; PARISE-MALTEMPI, 2020; SERRANO-FREITAS *et al.* 2020), demonstraram diferenças de distribuição entre os cromossomos de uma espécie e entre as espécies (Figura 6). Em relação à distribuição dos satélites nas espécies *C. gomesi* e *M. elongatus*, foi observada uma distinção na quantidade de cromossomos hibridados, a fêmea de *C. gomesi* apresentou uma maior quantidade de autossomos marcados por FISH em comparação à fêmea de *M. elongatus*.

A diferença encontrada no número de cromossomos hibridados entre espécies relacionadas também foi observada em outros trabalhos, sendo explicada pela hipótese das três etapas: nascimento, disseminação e agrupamento (RUIZ-RUANO *et al.*, 2016). Segundo a hipótese, os satDNAs podem seguir caminhos evolutivos distintos em diferentes espécies, proporcionando graus variáveis de propagação intercromossômica, ou seja, espécies relacionadas podem apresentar padrões de acúmulo e de distribuição distintas dos satDNAs devido à evolução independente desses elementos (CAMACHO *et al.*, 2021; RUIZ-RUANO *et al.*, 2016; UTSUNOMIA *et al.*, 2017).

Quanto à diferença encontrada à nível cromossômico dentro da própria espécie, o alto dinamismo relatado para as sequências de DNAs satélites pode justificar a variação dentro do cariótipo (MACAS *et al.*, 2010; MELTERS *et al.*, 2013; revisado por GARRIDO-RAMOS, 2015; PLOHL; MESTROVIC; MRAVINAC, 2012; ZHANG *et al.*, 2013; UTSUNOMIA *et al.*, 2017), no entanto ainda não foi apresentado com precisão quais mecanismos são responsáveis por essa variação (UTSUNOMIA *et al.*, 2017).

Além disso, e talvez ainda mais interessante de ser apontado, ambas espécies, *M. elongatus* e *C. gomesi*, também demonstraram distinção nos sinais de FISH em cromossomos específicos. *C. gomesi* apresentou marcações de Cgo44-177 no cromossomo B e nenhuma marcação nos cromossomos sexuais (Figura 6 A), diferentemente de *M. elongatus*, cujo cromossomo sexual heteromórfico  $W_1$  exibiu bandas visíveis de FISH para MelSat03-177 (Figura 6 B).

O satDNA MelSat03-177 corresponde a um dos satélites mais abundantes do genoma de *M. elongatus*, e junto de outros dois satélites mais abundantes representam 25% de todo o satelitoma diverso da espécie (CREPALDI *et al.*, 2021). Curiosamente, esse DNA satélite exibe grande acúmulo no  $W_1$ , mas não se encontra

enviesado para o sexo da fêmea, uma vez que apresenta taxas de abundância muito próximas entre os dois sexos. Os autores sugeriram que o acúmulo desse satélite no cromossomo sexual  $W_1$ , além dos autossomos, provavelmente tenha ocorrido ao acaso (CREPALDI *et al.*, 2021). Além disso, o satDNA MmaSat047-177 de *M. macrocephalus*, que demonstrou maior porcentagem de identidade com MelSat03-177, também não se apresenta enviesado para a fêmea (UTSUNOMIA *et al.*, 2019); contudo não existe dado de localização cromossômica para este satélite, sendo interessante uma análise posterior para saber se ele também se encontra acumulado no cromossomo W da espécie.

O trabalho com o satDNA Cgo44-177 de Serrano-Freitas *et al.* (2020) apresentou uma possível derivação do cromossomo B dos cromossomos sexuais na espécie *C. gomesi*, uma vez que a grande maioria dos satDNAs isolados mostraram-se específicos dos cromossomos sexuais da espécie. Essa derivação dos cromossomos supranumerários do complemento do cromossomo A já tem sido relatada em alguns trabalhos (revisado por GARRIDO-RAMOS, 2015; MARTIS *et al.*, 2012; JAMILENA *et al.*, 1995). No entanto, assim como foi visto para Cgo44-177, por mais que ocorra essa possível derivação, os cromossomos B podem apresentar um grau de acúmulo maior de DNAs satélites que também se encontram compartilhados nos cromossomos A, demonstrando que os supranumerários podem seguir caminhos evolutivos distintos do complemento A (revisado por GARRIDO-RAMOS, 2015; KLEMME *et al.*, 2013).

Com base nessas evidências, julgamos a necessidade de realizar análises futuras de mapeamento cromossômico dos demais satDNAs aqui estudados, a fim de verificar como essas sequências estão se comportando e acumulando nas outras espécies, principalmente para aquelas que possuem cromossomo B e/ou sistema de cromossomos sexuais heteromórficos; e de entender esse padrão atípico de acúmulo em alguns cromossomos autossomos e nos cromossomos específicos.

### 4.3 Função centromérica dos satDNAs estudados

A abundância elevada dos satDNAs no genoma pode apresentar indícios de funcionalidade centromérica (MELTERS *et al.*, 2013), à vista disso sabendo que MelSat03-177 é o terceiro satDNA mais abundante, representando 0.338% do genoma de *M. elongatus* (CREPALDI *et al.*, 2021), e que possui localização

centromérica, acredita-se que essas análises preliminares possam indicar uma possível função centromérica do satDNA MeISat03-177. No caso do satDNA Cgo44-177, apesar de não ser um dos DNAs satélites mais abundantes do genoma de *C. gomesi*, ele se encontra clusterizado na região centromérica em mais de 50% dos cromossomos da espécie, indicando também uma possível função centromérica. Por outro lado, Camacho *et al.* (2021) evidenciaram que as famílias de satDNA de uma superfamília que apresentam função centromérica em uma espécie não permanecem necessariamente com a mesma funcionalidade em espécies evolutivamente próximas.

Os estudos sobre os centrômeros no grupo dos peixes, principalmente no que concerne à presença da proteína CENP-B e a sua relação com os satDNAs centroméricos, ainda são incipientes. Por esse motivo, não sabemos se a ausência da caixa CENP-B nas sequências estudadas pode denotar a não funcionalidade centromérica destes satDNAs. À exemplo disso, Canapa *et al.* (2002) demonstraram um satDNA centromérico isolado da espécie *Gobius cobitis* que não apresenta homologia significativa com a caixa CENP-B, mas apresenta uma alta homologia com a sequência centromérica CDE III de levedura.

Apesar dos indícios da funcionalidade centromérica de MeISat03-177 e CgoSat44-177, a localização e o acúmulo dessas sequências presente nos centrômeros de alguns cromossomos é um fato incomum que coloca em questão a função centromérica desses satDNAs. Em geral, satDNAs que evidenciam funcionalidade centromérica tendem a se acumular em todos os cromossomos da espécie, como foi observado para outros grupos de organismos (DA SILVA *et al.*, 2020; MATSUDA; CHAPMAN, 1991; PAVLEK *et al.*, 2015; WILLARD, 1985; WU; MANUELIDIS, 1980).

Além disso, a condição para MeISat03-177 de não demonstrar um viés para a fêmea, mas apresentar certa importância na diferenciação do cromossomo sexual  $W_1$  (CREPALDI *et al.*, 2021) e a presença de CgoSat44-177 no cromossomo supranumerário (B), também coloca em questão a relevância dessas sequências repetitivas para esses cromossomos específicos, além da funcionalidade centromérica proposta inicialmente.

Desse modo, a diferença de arranjos desses satDNAs em todos os centrômeros dentro do cariótipo de *M. elongatus* e *C. gomesi* e a presença em

cromossomos sexuais e supranumerários, despertam questionamentos quanto a homogeneização e evolução dessas sequências dentro dessas espécies.

## **5 CONCLUSÃO**

Os resultados obtidos para as sequências recuperadas no presente trabalho condizem com a hipótese da biblioteca e a evolução independente propostas para DNAs satélites. Por meio de nossas análises foram demonstrados indícios de uma superfamília de DNAs satélites conservados em três famílias da Ordem Characiformes, bem como a sua presença em uma espécie da Ordem Gymnotiformes. Também foi apresentada uma possível funcionalidade centromérica para dois satDNAs recuperados e uma distribuição incomum destas sequências nos cromossomos de duas espécies analisadas.

A fim de entender melhor a conservação e a evolução dos DNAs satélites estudados, pretendemos futuramente realizar análises mais acuradas de bioinformática e localização cromossômica para os satDNAs recuperados.



## REFERÊNCIAS

ALBERTS, B. **Biologia Molecular da Célula**. 6 ed. São Paulo: Artmed, 2017.

BRITTEN, R. J.; KOHNE, D. E. Repeated sequences in DNA: Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. **Science**, v. 161, n. 3841, p. 529–540, 1968.

CAMACHO, J. P. M.; CABRERO, J.; LÓPEZ-LEÓN, M. D.; MARTÍN-PECIÑA, M. *et al.* On the contingent nature of satellite DNA evolution. **bioRxiv**, p. 2021.2007.2001.450629, 2021.

CANAPA, A.; CERIONI, P. N.; BARUCCA, M.; OLMO, E. *et al.* A centromeric satellite DNA may be involved in heterochromatin compactness in gobiid fishes. **Chromosome Research**, v. 10, n. 4, p. 297-304, 2002.

CARSON, R. L. **Silent Spring**. 1 ed. Boston: Houghton Mifflin, 1962.

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.

CREPALDI, C.; MARTÍ, E.; GONÇALVES, É. M.; MARTÍ, D. A. *et al.* Genomic Differences Between the Sexes in a Fish Species Seen Through Satellite DNAs. **Frontiers in Genetics**, 12, n. 1885, 2021.

DA SILVA, M. J.; DESTRO, R. F.; GAZONI, T.; NARIMATSU, H. *et al.* Great Abundance of Satellite DNA in *Proceratophrys* (Anura, Odontophrynidae) Revealed by Genome Sequencing. **Cytogenetic and Genome Research**, v. 160, p. 141-147, 2020.

DE BARROS, L. C. **Isolamento, caracterização e mapeamento cromossômico de elementos repetitivos em espécies da família Anostomidae (Ostariophysi – Characiformes): análise comparativa em diferentes tipos de águas amazônicas**. 2017. Tese (Doutorado em Genética, Conservação e Biologia Evolutiva) - Instituto Nacional de Pesquisas da Amazônia, Manaus, 2017.

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 BORBA, R. S.; DA SILVA, E. L.; PARISE-MALTEMPI, P. P. Chromosome mapping of retrotransposable elements. **Mobile Genetic Elements**, v. 3, n. 6, p. e27460, 2013.

DOS SANTOS, R. Z.; CALEGARI, R. M.; SILVA, D. M. Z. D. A.; RUIZ-RUANO, F. J. *et al.* 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. Molecular drive: a cohesive mode of species evolution. **Nature**, v. 299, n. 5879, p. 111-117, 1982.

DULZ, T. A.; AZAMBUJA, M.; NASCIMENTO, V. D.; LORSCHIEDER, C. A. *et al.* Karyotypic Diversification in Two *Megaleporinus* Species (Characiformes, Anostomidae) Inferred from In Situ Localization of Repetitive DNA Sequences. **Zebrafish**, n. 5, p. 333-341, 2020.

DUMONT M.; FACHINETTI, D. DNA Sequences in Centromere Formation and Function. **Progress in Molecular and Subcellular Biology**, v. 56, p. 305-336, 2017.

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

EDWARDS, N. S.; MURRAY, A. W. Identification of *Xenopus* CENP-A and an Associated Centromeric DNA Repeat. **Molecular Biology of the Cell**, v. 16, n. 4, p. 1800-1810, 2005.

ELDER JR, J. F.; TURNER, B. J. Concerted Evolution of Repetitive DNA Sequences in Eukaryotes. **The Quarterly Review of Biology**, v. 70, n. 3, 1995.

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

FUJITA, R.; OTAKE, K.; ARIMURA, Y.; HORIKOSHI, N. *et al.* Stable complex formation of CENP-B with the CENP-A nucleosome. **Nucleic Acids Research**, v. 43, n. 10, p. 4909-4922, 2015.

GALETTI, J. P. M.; 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.

GALETTI JR, P. M.; LIMA, N. R. W.; VENERE, P. C. A Monophyletic ZW Sex Chromosome System in *Leporinus* (Anostomidae, Characiformes). **Cytologia**, v. 60, n. 4, p. 375-382, 1995.

GALETTI, P. M.; CESAR, A. C. G.; VENERE, P. C. Heterochromatin and NORs variability in *Leporinus* fish (Anostomidae, Characiformes). **Caryologia**, v. 44, n. 3-4, p. 287-292, 1991.

GARRIDO-RAMOS, M. A. Satellite DNA: An evolving topic. **Genes**, v. 8, n. 9, p. 230-271, 2017.

GARRIDO-RAMOS, M. A. Satellite DNA in Plants: More than Just Rubbish. **Cytogenetic and Genome Research**, v. 146, n. 2, p. 153-170, 2015.

HASHIMOTO, D. T.; PARISE-MALTEMPI, P. P.; LAUDICINA, A.; BORTOLOZZI, J. *et al.* 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.

HENIKOFF, S.; AHMAD, K.; MALIK, H. S. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. **Science**, v. 293, n. 5532, p. 1098–1102, 2001.

HERRÁN *et al.* Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of Sturgeons. **Molecular Biology and Evolution**, v. 18, n. 3, p. 432-436, 2001.

HERRÁN *et al.* The centromeric satellite of the wedge sole (*Dicologlossa cuneata*, Pleuronectiformes) is composed mainly of a sequence motif conserved in other vertebrate centromeric DNAs. **Cytogenetic and Genome Research**, v. 121, n. 3-4, p. 271–276, 2008.

JAMILENA, M.; GARRIDO-RAMOS, M. A.; RUIZ REJÓN, M.; RUIZ REJÓN, C. *et al.* Characterisation of repeated sequences from microdissected B chromosomes of *Crepis capillaris*. **Chromosoma**, v. 104, n. 2, p. 113-120, 1995.

KIT, S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. **Journal of Molecular Biology**, v. 3, p. 711-716, 1961.

KLEMME, S.; BANAEI-MOGHADDAM, A. M.; MACAS, J.; WICKER, T. *et al.* High-copy sequences reveal distinct evolution of the rye B chromosome. **New Phytologist**, v. 199, n. 2, p. 550-558, 2013.

KURSEL, L. E.; MALIK, H. S. The cellular mechanisms and consequences of centromere drive. **Current Opinion in Cell Biology**, v. 52, p. 58-65, 2018.

LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. A. The repetitive DNA content of eukaryotic genomes. **Genome Dynamics**, v. 7, p. 1-28, 2012.

LORITE, P.; MUÑOZ-LÓPEZ, M.; CARRILLO, J. A.; SANLLORENTE, O. *et al.* Concerted evolution, a slow process for ant satellite DNA: study of the satellite DNA

in the *Aphaenogaster* genus (Hymenoptera, Formicidae). **Organisms Diversity and Evolution**, v.17, p. 595-606, 2017.

LOUZADA, S.; LOPES, M.; FERREIRA, D.; ADEGA, F. *et al.* Decoding the Role of Satellite DNA in Genome Architecture and Plasticity—An Evolutionary and Clinical Affair. **Genes**, v. 11, n. 1, p. 72, 2020.

MACAS, J.; NEUMANN, P.; NOVÁK, P.; JIANG, J. Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. **Bioinformatics**, v. 26, n. 17, p. 2101-2108, 2010.

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, p. 363–367, 1999.

MARTIS, M. M.; KLEMME, S.; BANAEI-MOGHADDAM, A. M.; BLATTNER, F. R. *et al.* Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 33, p. 13343-13346, 2012.

MATSUDA, Y.; CHAPMAN, V. M. In situ analysis of centromeric satellite DNA segregating in *Mus* species crosses. **Mammalian Genome**, v. 1, n. 71, 1991.

MELTERS, D. P.; BRADNAM, K. R.; YOUNG, H. A.; TELIS, N. *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. **Genome Biology**, v. 14, n. 1, p. R10, 2013.

MESTROVIC, N.; PLOHL, M.; MRAVINAC, B.; UGARKOVIC, D. Evolution of satellite DNAs from the genus *Palorus*--experimental evidence for the "library" hypothesis. **Molecular Biology and Evolution**, v. 15, n. 8, p. 1062-1068, 1998.

MESTROVIC, N.; PAVLEK, M.; CAR, A.; CASTAGNONE-SERENO, P. *et al.* Conserved DNA Motifs, Including the CENP-B Box-like, Are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. **PLOS ONE**, v. 8, n. 6, p. e67328, 2013.

METZNER, R. Beitrage zur Granulalehre. **Archiv für Anatomie und Physiologie**, p. 309, 1894.

MOLINA, W. F.; GALETTI, P. M. Early replication banding in *Leporinus* species (Osteichthyes, Characiformes) bearing differentiated sex chromosomes (ZW). **Genetica**, v. 130, n. 2, p. 153-160, 2007.

NASCIMENTO, M.; SOUSA, A.; RAMIREZ, M.; FRANCISCO, A. P. *et al.* PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. **Bioinformatics**, v. 33, n. 1, p. 128-129, 2017.

NOVÁK, P.; NEUMANN, P.; PECH, J.; STEINHAIŠL, J.; MACAS, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 1, p.1-2, 2013.

PALACIOS-GIMENEZ, O. M.; MILANI, D.; SONG, H.; MARTI, D. A. *et al.* Eight Million Years of Satellite DNA Evolution in Grasshoppers of the Genus *Schistocerca* Illuminate the Ins and Outs of the Library Hypothesis. **Genome Biology and Evolution**, v. 12, n. 3, p. 88-102, 2020.

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.

PAVLEK, M.; GELFAND, Y.; PLOHL, M.; MESTROVIC, N. Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. **DNA Research**, v. 22, n. 6, p. 387-401, 2015.

PERROTTI, P. B.; ABDO, T. F.; MEIRELES, W. A.; MARCON, L. *et al.* Ontogenesis and embryogenesis of *Megaleporinus elongatus* larvae (Characiformes: Anostomidae) from the Jequitinhonha River Basin. **Journal of Applied Ichthyology**, v. 35, n. 2, p. 520-528, 2018.

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, p. 72-82, 2008.

PLOHL, M.; MESTROVIC, N.; MRAVINAC, B. Satellite DNA evolution. **Genome Dynamics**, v. 7, p. 126-152, 2012.

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

POLTRONIERI, J.; MARQUIONI, V.; BERTOLLO, L. A. C.; KEJNOVSKY, E. *et al.* 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.

RAMIREZ, J. L.; BIRINDELLI, J. L.; CARVALHO, D. C.; AFFONSO, P. R. A. M. *et al.* Revealing Hidden Diversity of the Underestimated Neotropical Ichthyofauna: DNA Barcoding in the Recently Described Genus. **Frontiers in Genetics**, v. 8, p. 149, 2017.

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, 2017.

REIS, R. E.; KULLANDER, S. O.; FERRARIS JR, C. J. **Check list of the freshwater fishes of South and Central America**. [s.n.]. Porto Alegre: EDIPUCRS, 2003. Disponível em: <https://bitly.com/kBihVf>. Acesso em: 20 mar. 2021.

RHOADES, M. M. Studies of a telocentric chromosome in maize with reference to the stability of its centromere. **Genetics**, v. 25, n. 5, p. 483, 1940.

RODRÍGUEZ, F. R. et al. Centromeric satellite DNA in flatfish (Order Pleuronectiformes) and its relation to speciation processes. **Journal of Heredity**, p. 217-222, 2017.

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.; CASTILLO-MARTÍNEZ, J.; CABRERO, J.; GÓMEZ, R. *et al.* High-throughput analysis of satellite DNA in the grasshopper *Pyrgomorpha conica* reveals abundance of homologous and heterologous higher-order repeats. **Chromosoma**, v. 127, n. 3, p. 323-340, 2018.

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, n. 1, p. 28333, 2016.

SERRANO-FREITAS, É. A.; SILVA, D. M. Z. A.; RUIZ-RUANO, F. J.; UTSUNOMIA, R. *et al.* Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. **Molecular Genetics and Genomics**, v. 295, n. 1, p. 195-207, 2020.

SIDLAUSKAS, B. L.; ASSEGA, F. M.; MELO, B. F.; OLIVEIRA, C. *et al.* Total evidence phylogenetic analysis reveals polyphyly of Anostomoides and uncovers an unexpectedly ancient genus of Anostomidae fishes (Characiformes). **Zoological Journal of the Linnean Society**, 2021. <https://doi.org/10.1093/zoolinnean/zlab016>.

SIDLAUSKAS, B. L.; VARI, R. P. Phylogenetic relationships within the South American fish family Anostomidae (Teleostei, Ostariophysi, Characiformes). **Zoological Journal of the Linnean Society**, v. 154, n. 1, p. 70-210, 2008.

SUEOKA, N. Variation and heterogeneity of base composition of deoxyribonucleic acids: A compilation of old and new data. **Journal of Molecular Biology**, v. 3, n. 1, p. 31-IN15, 1961.

TEIXEIRA, J. R.; DIAS, G. B.; SVARTMAN, M.; RUIZ, A. *et al.* Concurrent duplication of the Cid and Cenp-C genes in the *Drosophila* subgenus with signatures of

subfunctionalization and male germline-biased expression. **bioRxiv**, p. 134817, 2018.

THAKUR, J.; PACKIARAJ, J.; HENIKOFF, S. Sequence, Chromatin and Evolution of Satellite DNA. **International Journal of Molecular Sciences**, v. 22, n. 9, p. 4309, 2021.

UGARKOVIC, D.; PLOHL, M. Variation in satellite DNA profiles—causes and effects. **The EMBO Journal**, v. 21, n. 22, p. 5955-5959, 2002.

UTSUNOMIA, R.; RUIZ-RUANO, F. J.; SILVA, D. M. Z. A.; SERRANO, É. A. *et al.* A Glimpse into the Satellite DNA Library in Characidae Fish (Teleostei, Characiformes). **Frontiers in Genetics**, v. 8, n. 103, 2017.

UTSUNOMIA, R.; SILVA, D. M. Z. D. A.; RUIZ-RUANO, F. J.; GOES, C. A. G. *et al.* Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. **Scientific Reports**, 9, n. 1, p. 5856, 2019.

WILLARD, H. F. Chromosome-specific organization of human  $\alpha$  satellite DNA. **American Journal of Human Genetics**, v. 37, n. 3, p. 524-532, 1985.

WU, J. C.; MANUELIDIS, L. Sequence definition and organization of a human repeated DNA. **Journal of Molecular Biology**, v. 142, n. 3, p. 363-386, 1980.

YU, F.; DOU, Q.; LIU, R.; WANG, H. A conserved repetitive DNA element located in the centromeres of chromosomes in *Medicago* genus. **Genes and Genomics**, v. 39, p. 903-911, 2017.

ZHANG, T.; TALBERT, P. B.; ZHANG, W.; WU, Y. *et al.* The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. **PNAS December**, v. 110, n. 50, p. E4875-E4883, 2013.