



Universidade Estadual Paulista (UNESP)
Faculdade de Ciências e Letras de Assis

**Um novo conjunto de DNAs satélites
compartilhado por todos os Characoidei**

(Characiformes, Teleostei)

Dissertação de Mestrado

**Leonardo Moura Giglio
Bauru, 2022**

LEONARDO MOURA GIGLIO

**Um novo conjunto de DNAs satélites compartilhado por todos os
Characoidei (Characiformes, Teleostei)**

**BAURU
2022**

LEONARDO MOURA GIGLIO

**Um novo conjunto de DNAs satélites compartilhado por todos os
Characoidei (Characiformes, Teleostei)**

Dissertação apresentada à Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras de Assis, para a obtenção do título de Mestre em Biociências (Área do conhecimento: Caracterização e Aplicação da Diversidade Biológica).

Orientador: Dr. Ricardo Utsunomia

Co-Orientador: Dr. Fábio Porto-Foresti

Bolsista: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 88887.483868/2020-00

BAURU

2022

G459n Giglio, Leonardo Moura
Um novo conjunto de DNAs satélites compartilhado por
todos os Characoidei (Characiformes, Teleostei) /
Leonardo Moura Giglio. -- Assis, 2022
56 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista
(Unesp), Faculdade de Ciências e Letras, Assis
Orientador: Ricardo Utsunomia
Coorientador: Fábio Porto-Foresti

1. Genética Animal. 2. Genômica. 3. DNA repetitivo. 4.
Ostariophysi. 5. Bioinformática. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências e Letras, Assis. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE LEONARDO MOURA GIGLIO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCIÊNCIAS, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.

Aos 01 dias do mês de julho do ano de 2022, às 09:00 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de LEONARDO MOURA GIGLIO, intitulada **Um novo conjunto de DNAs satélites compartilhado por todos os Characoidei (Characiformes, Teleostei)**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. RICARDO UTSUNOMIA (Participação Virtual) do(a) UFRRJ/Rio de Janeiro, Prof. Dr. FRANCISCO JESÚS RUIZ-RUANO CAMPAÑA (Participação Virtual) do(a) University of East Anglia/UK, Profa. Dra. MARA CRISTINA DE ALMEIDA (Participação Virtual) do(a) Departamento de Biologia Estrutural, Molecular e Genética / Universidade Estadual de Ponta Grossa. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final: aprovado. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.



Prof. Dr. RICARDO UTSUNOMIA

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 88887.483868/2020-00.

Gostaria de agradecer primeiramente à minha família, mãe, pai, irmão e avós, pelo suporte financeiro e emocional por todos estes longos anos de caminhada acadêmica, sem o qual não estaria aqui; e à minha parceira de vida, Mirela, por estar sempre ao meu lado.

Gostaria também de agradecer a todos meus amigos de São José do Rio Preto e de Bauru, de toda a universidade, mas especialmente aos amigos da república, que compartilharam todos estes anos comigo, e aos amigos do laboratório, pelo grande companheirismo e por todas as risadas.

Por último, gostaria de agradecer aos professores e orientadores, Ricardo e Fábio, pela grande oportunidade cedida.

GIGLIO, Leonardo Moura. **Um novo conjunto de DNAs satélites compartilhado por todos os Characoidei (Characiformes, Teleostei)**. 2022. 56 f. Dissertação (Mestrado Acadêmico em Biociências) – Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Assis, 2022.

RESUMO

O genoma dos organismos eucariotos apresenta uma porção repetitiva bastante expressiva, com a classe dos DNAs satélites podendo representar grande parte do DNA nuclear e, também, contribuir para a diferença do tamanho dos genomas entre diferentes espécies. Considera-se que os DNAs satélites surgem a partir de duplicações *de novo* de sequências preexistentes no genoma das espécies, se disseminam, e posteriormente expandem seus arranjos, a forma em que se organizam. Esses arranjos passam por processos de homogeneização e posterior fixação de sua composição, sob o regime da evolução em concerto. Esses mecanismos evolutivos podem agir independentemente nas diferentes espécies, inclusive, agir independentemente nos diferentes arranjos dentro de um mesmo genoma, o que normalmente resulta em DNAs satélites gênero ou espécie específicos. Porém, hoje temos conhecimento de alguns casos de satélites que se mantiveram conservados em diferentes espécies distribuídas por táxons maiores, como toda uma ordem, apresentando inclusive transcrição, o que vai contra as previsões da evolução em concerto em um contexto de ausência de função. Outro modelo de evolução de DNAs satélites é o da hipótese *library*, que prediz que grupos proximalmente relacionados compartilham uma porção do conteúdo ancestral de DNAs satélites, apresentando normalmente diferenças quantitativas relativas a sua abundância no genoma. No presente estudo utilizamos a subordem dos peixes Characoidei, da ordem dos Characiformes, altamente diversa e com origem antiga, e portanto considerada ideal para verificar a existência de satélites compartilhados pelo grupo, testar a hipótese *library*, e analisar a ação da evolução em concerto nos satélites destas espécies. Para isto, utilizamos bibliotecas genômicas de 15 espécies de diferentes famílias da ordem dos Characiformes e realizamos análises bioinformáticas com a aplicação do TAREAN, uma ferramenta que utiliza um algoritmo de agrupamento de *reads* para realizar o isolamento dos DNAs satélites das espécies, além de ferramentas de montagem *de novo*, alinhamento e mapeamento de cópias para realizar a comparação destes satélites entre as espécies. Encontramos sete DNAs satélites presentes nas quatro superfamílias de peixes Characoidei analisadas, demonstrando o compartilhamento de satélites super conservados por toda uma subordem, o que contraria as previsões da evolução em concerto, e também corroboram com a hipótese *library* de evolução dos DNAs satélites.

Palavras-chave: DNA repetitivo. Ostariophysi. Bioinformática.

GIGLIO, Leonardo Moura. **A new satellite DNA set shared by all Characoidei (Characiformes, Teleostei)**. 2022. 56 p. Dissertation (Masters in Bioscience) – Sao Paulo State University (UNESP), School of Science, Humanities and Languages, Assis, 2022.

ABSTRACT

The eukaryotic genomes present a very expressive repetitive portion, with the satellite DNAs class representing a great part of the nuclear DNA and, also, contributing to the differences observed in the species genome sizes. It is assumed that the satellite DNAs originate by de novo duplications in preexistent sequences in the genome, are scattered to different locations and, subsequently, expand their arrays, the way they are organized. These arrays go through composition homogenization and fixation steps, under the concerted evolution regime. These evolutive mechanisms can act independently in each species, or even act differently in arrays of different genomic locations, and that usually results in genus- or species-specific satellite DNAs. However, today we know of some cases of satellites that have remained conserved in different species distributed over larger taxa, such as an entire order, even presenting transcription, which goes against the predictions of concerted evolution in a context of absent function. Another model of satellite DNA evolution is the library hypothesis, which predicts that closely related groups share a portion of the ancestral content of satellite DNAs, typically showing quantitative differences relative to their abundance in the genome. In the present study, we used the suborder of Characoidei fishes, from the order Characiformes, highly diverse and with ancient origin, and therefore considered ideal to verify the existence of satellites shared by the group, test the library hypothesis, and analyze the action of the concerted evolution in the satellites of these species. For this, we used genomic libraries of 15 species from different families of the Characiformes order and performed bioinformatic analyses with the application of TAREAN, a tool that uses a read clustering algorithm to isolate the species satellite DNAs, as well as de novo assembly, alignment, and copy mapping tools to compare these satellites between species. We found seven satellite DNAs present in the four analyzed Characoidei fish superfamilies, demonstrating the sharing of super conserved satellites across an entire suborder, which contradicts the predictions of concerted evolution, and also corroborates with the library hypothesis of satellite DNA evolution.

Keywords: repetitive DNA. Ostariophysi. Bioinformatics.

SUMÁRIO

1. INTRODUÇÃO	7
2. OBJETIVOS	10
3. MATERIAL E MÉTODOS.....	11
3.1 Obtenção das amostras e extração de DNA genômico.....	13
3.2 Análises bioinformáticas.....	14
3.2.1 Isolamento de DNAs satélites com o TAREAN.....	14
3.2.2 Caracterização do satelitoma das espécies.....	16
3.2.3 Comparações entre os catálogos das espécies.....	17
4. RESULTADOS.....	18
4.1 Descrição de DNAs satélites nas diferentes espécies.....	18
4.2 Análises comparativas dos catálogos de satélites.....	25
5. DISCUSSÃO.....	34
5.1 Caracterização de satelitomas e geração de recursos genômicos.....	34
5.2 Evolução e compartilhamento de satélites conservados.....	36
5.3 Considerações finais.....	39
6. REFERÊNCIAS.....	40
7. MATERIAL SUPLEMENTAR.....	45

1. INTRODUÇÃO

O genoma dos eucariotos pode ser subclassificado em uma porção gênica, que compreende todos os genes de cópia única, e por uma porção repetitiva do DNA, que compreende diferentes tipos de sequências, como as famílias multigênicas (com genes que apresentam muitas cópias no genoma), os elementos transponíveis dispersos pelo genoma, os DNAs satélites (satDNAs) formando arranjos (*arrays*) organizados de seus monômeros, os pseudogenes, derivados de genes funcionais, e os RNAs não codificantes (ncRNA), podendo apresentar diferentes funções celulares. O DNA repetitivo representa uma porção substancial do genoma, compreendendo mais da metade do DNA nuclear em grande parte das espécies e, conseqüentemente, é o principal responsável pelas diferenças dos tamanhos dos genomas de diferentes organismos (CHARLESWORTH; SNIEGOWSKI; STEPHAN, 1994; AMARAL; MATTICK, 2008; TUTAR, 2012; BISCOTTI; OLMO; HESLOP-HARRISON, 2015).

Os DNAs satélites formam uma classe de sequências que se encontram repetidas formando arranjos de suas subunidades organizadas em tandem (repetidas sequencialmente em uma orientação “cabeça-cauda”), que podem variar de poucos a milhares de nucleotídeos (DOS SANTOS et al., 2021). O conjunto de satélites de uma espécie (satelitoma) pode representar uma porção significativa de seu genoma, como no caso do gafanhoto *Locusta migratoria*, com um conteúdo de satélites que compreende aproximadamente 2,4% de seu genoma, e do peixe *Megaleporinus macrocephalus*, com aproximadamente 12% de seu genoma composto por satélites (PLOHL; MEŠTROVIĆ; MRAVINAC, 2012; RUIZ-RUANO et al., 2016; GARRIDO-RAMOS, 2017; UTSUNOMIA et al., 2019). Os satDNAs estão localizados principalmente na heterocromatina, nas regiões pericentromérica e subtelomérica, mas não são restritos à elas (PLOHL; MEŠTROVIĆ; MRAVINAC, 2012; GARRIDO-RAMOS, 2015; KUHN, 2015; RUIZ-RUANO et al., 2016; SILVA et al., 2017; MIGA et al., 2020).

A origem de novos DNAs satélites geralmente se dá a partir da duplicação de um fragmento de sequência preexistente, que pode então ser disperso pelo genoma e localmente amplificado por meio de translocação, do deslizamento da DNA polimerase durante o processo de replicação, de inserções, ou por *crossing-over* desigual, também podendo ser extinto a partir de processos estocásticos (PLOHL; MEŠTROVIĆ; MRAVINAC, 2012; RUIZ-RUANO et al., 2016; LOWER et al., 2018; VONDRAK et al., 2020). Considera-se que os satDNAs evoluem por meio de evolução em concerto, que compreende uma fase de homogeneização dos arranjos de determinada sequência conduzida por mecanismos

moleculares como *crossing-over* desigual e conversão gênica, e uma fase de fixação que ocorre naturalmente com a reprodução sexuada dos organismos. Como consequência deste processo, satDNAs usualmente apresentam uma alta homogeneidade intraespecífica e diversificação interespecífica, levando muitas destas sequências a serem espécie- ou gênero-específicas (SMITH, 1976; DOVER, 1982, 1986; PLOHL; MEŠTROVIĆ; MRAVINAC, 2012; GARRIDO-RAMOS, 2017; LORITE et al., 2017; HALBACH et al., 2020). Em 1977, Fry e Salser propuseram o modelo da hipótese *library* (biblioteca, em uma tradução livre) de evolução de DNA satélites, que afirma que espécies proximamente relacionadas compartilham uma porção do conteúdo destas sequências no genoma, mas exibem, após um longo prazo evolutivo, um número de eventos de expansão de número de cópias independente e distinto em cada uma das linhagens, o que acaba resultando em diferentes padrões de amplificação ou de diminuição dos arranjos de cada satélite nas diferentes espécies, produzindo as diferenças quantitativas existentes entre suas bibliotecas de DNAs satélites.

Por serem elementos genômicos com uma dinâmica evolutiva bem complexa e não muito bem compreendida (BISCOTTI; OLMO; HESLOP-HARRISON, 2015), o estudo dos DNAs satélites foi bastante dificultado no passado. Desde a década de 70, as principais metodologias utilizadas para a identificação de novas sequências deste tipo eram a digestão com enzimas de restrição e/ou a cinética de renaturação (HUTTON; WETMUR, 1973; BOTCHAN, 1974; HÖRZ; ZACHAU, 1977; KOPECKA et al., 1978), metodologias pouco efetivas e que, no caso da restrição enzimática, dependente da clivagem por acaso em sítios específicos de algum satélite desconhecido. No entanto, este cenário se alterou com o passar dos anos por conta do aperfeiçoamento das tecnologias de sequenciamento massivo paralelo, bem como pelo desenvolvimento de ferramentas bioinformáticas adequadas que permitissem a análise da grande quantidade de dados gerada. Em 2013, por exemplo, o portal Galaxy, da infraestrutura internacional Elixir, voltada para o estudo da biologia, implementou o conjunto de ferramentas para a análise de sequências repetitivas RepeatExplorer (NOVÁK et al., 2013), com um método inovador de identificação *de novo* de sequências repetitivas a partir de *reads* curtos de sequenciamento genômico, baseado no agrupamento dos *reads* em *clusters*, e na inspeção visual dos gráficos resultantes, que em conjunto com o protocolo satMiner, desenvolvido por Ruiz-Ruano et al. (2016), permitiu o início da caracterização dos satelitomas das espécies. Em 2017, uma nova ferramenta foi implementada, o TAREAN (*Tandem Repeat Analyzer* - NOVÁK et al., 2017), automatizando a análise dos gráficos gerados e facilitando a caracterização de catálogos robustos de DNAs satélites, que levou ao aumento expressivo no número de satélites descritos em espécies não-modelo (RUIZ-

RUANO et al., 2016; SILVA et al., 2017; UTSUNOMIA et al., 2019; CREPALDI; PARISE-MALTEMPI, 2020; SERRANO-FREITAS et al., 2020; STORNIOLI et al. 2021; CAMACHO et al., 2022; KRETSCHMER et al., 2022).

A ordem dos Characiformes (superordem Ostariophysi) é um dos clados de peixes mais ricos em número de espécies do mundo, representando grande parte da diversidade de peixes da ictiofauna Neotropical (MOREIRA, 2007; ARCILA et al., 2017; BETANCUR-R et al., 2019; FRICKE, 2021). A origem antiga da ordem é bem estabelecida, com dados moleculares que apontam para a existência do último ancestral em comum dos peixes da superordem há aproximadamente 150 milhões de anos atrás (NEAR et al., 2012; DAI et al., 2018; BURNS; SIDLAUSKAS, 2019; MELO et al., 2022), e recentemente, Melo et al. (2022) reforçaram a ideia da parafilia dos Characiformes, com os Siluriformes como grupo irmão de Characoidei, ambos formando um clado irmão de Citharinoidei, embora ainda não exista um consenso sobre as relações filogenéticas recalcitrantes nestes grupos (FINK; FINK, 1981; DIMMICK; LARSON, 1996; NAKATANI et al., 2011; CHEN; LAVOUÉ; MAYDEN, 2013; BETANCUR-R et al., 2013; ARCILA et al., 2017; CHAKRABARTY et al., 2017).

Uma outra questão interessante neste grupo é a origem do alto número de espécies e eventos de radiação adaptativa observados em determinadas linhagens (MELO et al., 2022). Consequentemente, essa alta diversidade de espécies do grupo também levanta diversas outras problemáticas para serem estudadas, como a presença, origem e diferenciação de cromossomos sexuais e supranumerários (SILVA et al., 2017; UTSUNOMIA et al., 2019; CREPALDI; PARISE-MALTEMPI, 2020; SERRANO-FREITAS et al., 2020; STORNIOLI et al., 2021; KRETSCHMER et al., 2022). Neste contexto, consideramos os peixes Characiformes ideais para o estudo da dinâmica evolutiva dos DNAs satélites.

Até recentemente, menos de uma dezena de satélites havia sido caracterizada para peixes Neotropicais (revisão em VICARI et al., 2010). Desde então, a caracterização de satelitomas tem permitido uma melhor compreensão sobre os mecanismos que regem a evolução destas sequências, além de permitir sua aplicação na análise da distribuição dos DNAs satélites pelo genoma das espécies, resultando em importantes evidências para o entendimento de diferentes problemáticas evolutivas em peixes (SILVA et al., 2017; UTSUNOMIA et al., 2019; CREPALDI; PARISE-MALTEMPI, 2020; SERRANO-FREITAS et al., 2020; STORNIOLI et al., 2021; KRETSCHMER et al., 2022).

No que se refere ao compartilhamento de satDNAs entre espécies relacionadas, alguns poucos exemplos incluem satélites que se mantiveram conservados em alguns grupos por longos períodos de tempo: como a família do *Alpha* Satélite em primatas, com idade estimada

na ordem de 35 milhões de anos (WILLARD, 1991; ALEXANDROV et al., 2001); o DNA satélite APSP-I, similar em várias espécies do gênero *Arphaenogaster* de formigas, com idade estimada entre 80-74 milhões de anos (LORITE et al., 2017); de um satélite conservado por aproximadamente 200 milhões de anos no gênero *Aedes* que controla a expressão de alguns genes durante o desenvolvimento (HALBACH et al., 2020); e do caso que mais inspira esse trabalho, que é o CharSat01, um satélite presente em 14 espécies da superfamília de peixes Characoidei, e que é ativamente transcrito (DOS SANTOS et al., 2021). Deve-se destacar que casos de satélites compartilhados desafiam as predições de evolução molecular de satDNAs por evolução em concerto, o que levanta dúvidas quanto a possível existência de uma função genômica/celular destas sequências, o que já foi confirmado nos casos do *Alpha* Satélite, do APSP-I, e dos tapiR1 e tapiR2 em *Aedes* (ALEXANDROV et al., 2001; LORITE et al., 2017; SMALEC et al., 2019; DOS SANTOS et al., 2021).

Diante do exposto, consideramos que a caracterização de novos satelitomas de espécies pertencentes à subordem dos Characoidei com o uso de bibliotecas genômicas provenientes de sequenciamento de segunda geração, aliado à aplicação de diferentes ferramentas bioinformáticas, possibilitará um estudo comparativo entre peixes de diferentes famílias distribuídas pelo grupo, o que permitirá testar a hipótese *library* e verificar se existem mais satélites conservados por toda a subordem, além de analisar a atuação da evolução em concerto nos arranjos dos satélites. No presente estudo nós caracterizamos cinco novos satelitomas de peixes Characoidei, além de complementar um catálogo previamente descrito. Além disso, nossas análises evidenciaram a ocorrência de sete satélites compartilhados por diferentes famílias da subordem, expandindo o conhecimento da diversidade de satélites em peixes e servindo de subsídio para diversos estudos citogenéticos e de bioinformática no futuro.

2. OBJETIVOS

O objetivo geral deste trabalho foi identificar novos casos de satélites conservados ao longo de todos os Characoidei. Para isso, os objetivos específicos foram:

- A) Caracterizar 5 novos satelitomas das espécies pertencentes à diferentes famílias dos Characoidei: *Hoplias malabaricus*, *Leporinus friderici*, *Cyphocharax modestus*, *Pyrrhulina semifasciata* e *Astyanax mexicanus*, além de complementar o

satelitoma de *Characidium gomesi*, previamente descrito em Serrano-Freitas, 2020;

B) Identificar satélites compartilhados por *C. gomesi* e por outra espécie da subordem que apresentem alta conservação de sua composição.

3. MATERIAL E MÉTODOS

No presente estudo, foram utilizadas bibliotecas genômicas de 15 espécies de peixes da subordem Characoidei, representando 10 diferentes famílias, sendo *C. gomesi*, da família Crenuchidae, considerado o clado mais basal dentro do grupo. A relação de todas as espécies envolvidas no estudo e suas respectivas famílias está descrita na Tabela 1.

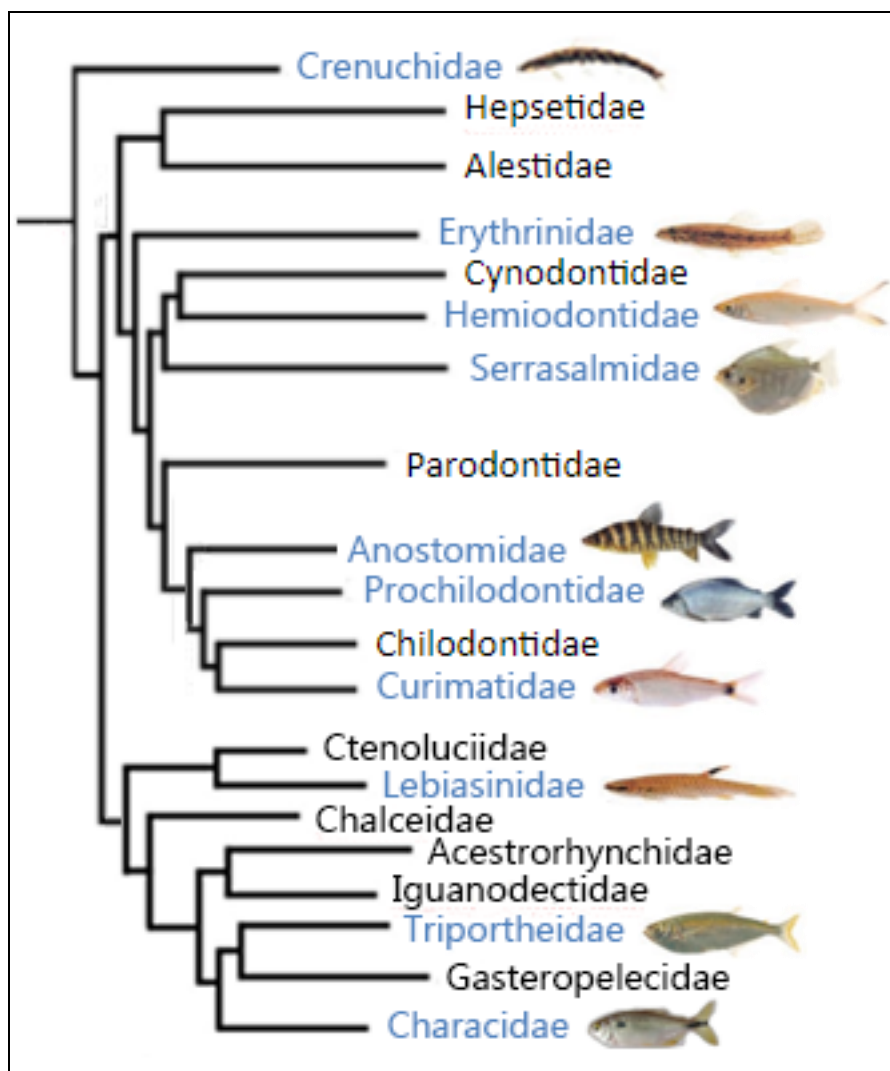


Figura 1: Famílias das espécies analisadas (em azul) e sua distribuição pela subordem Characoidei.

Tabela 1: Espécies utilizadas no projeto e suas respectivas famílias dentro da subordem dos Characoidei.

Espécie	Família
<i>Characidium gomesi</i>	Crenuchidae
<i>Hoplias malabaricus</i>	Erythrinidae
<i>Hemiodus gracilis</i>	Hemiodontidae
<i>Pygocentrus nattereri</i>	Serrasalminidae
<i>Myleus asterias</i>	Serrasalminidae
<i>Colossoma macropomum</i>	Serrasalminidae
<i>Piaractus brachypomus</i>	Serrasalminidae
<i>Piaractus mesopotamicus</i>	Serrasalminidae
<i>Leporinus friderici</i>	Anostomidae
<i>Prochilodus lineatus</i>	Prochilodontidae
<i>Cyphocharax modestus</i>	Curimatidae
<i>Pyrrhulina semifasciata</i>	Lebiasinidae
<i>Triportheus auritus</i>	Triportheidae
<i>Psalidodon paranae</i>	Characidae
<i>Astyanax mexicanus</i>	Characidae

As informações sobre as bibliotecas genômicas das espécies, incluindo a plataforma de sequenciamento e seu tamanho, podem ser observadas na Tabela 2. Para o presente trabalho, utilizamos algumas bibliotecas provenientes de trabalhos anteriores do nosso grupo de pesquisa ou de colaboradores, citados na tabela, além de uma biblioteca obtida do banco de dados *Sequence Read Archive*, do NCBI (*A. mexicanus*).

162 **Tabela 2:** Plataformas de sequenciamento e tamanho das bibliotecas utilizadas.

Espécie	Plataforma	Tamanho total (GB)	Acesso NCBI ou Referência
<i>C. gomesi</i>	Illumina HiSeq	0,8 GB	-
<i>H. malabaricus</i>	BGISEQ-500	1,4 GB	-
<i>H. gracilis</i>	Illumina MiSeq	0,5 GB	Rodrigues (2021)
<i>P. nattereri</i>	Illumina HiSeq	59 GB	Rodrigues (2021)
<i>M. asterias</i>	Illumina MiSeq	0,6 GB	Rodrigues (2021)
<i>C. macropomum</i>	MGISEQ-2000	1,6 GB	Rodrigues (2021)
<i>P. brachypomus</i>	MGISEQ-2000	2,7 GB	Rodrigues (2021)
<i>P. mesopotamicus</i>	Illumina MiSeq	0,4 GB	Rodrigues (2021)
<i>L. friderici</i>	Illumina MiSeq	0,2 GB	-
<i>P. lineatus</i>	BGISEQ-500	3,6 GB	Stornioli et al. (2021)
<i>C. modestus</i>	BGISEQ-500	2,4 GB	-
<i>P. semifasciata</i>	BGISEQ-500	2,1 GB	-
<i>T. auritus</i>	BGISEQ-500	3,4 GB	Kretschmer et al. (2022)
<i>P. paranae</i>	BGISEQ-500	17 GB	dos Santos, em elaboração.
<i>A. mexicanus</i>	Illumina HiSeq	54 GB	SRA - SRR6386652

163
 164 De um total de 15 espécies utilizadas no estudo (Tabela 2), nove delas já possuíam seu
 165 satelitoma caracterizado com o uso do TAREAN, portanto, as etapas 3.1, 3.2.1 e 3.2.2
 166 (abaixo) não foram realizadas para estas espécies, com exceção da análise de correlação
 167 estatística dos traços dos satelitomas (incluída na etapa de caracterização dos satelitomas),
 168 realizada com todos os catálogos.

170 3.1 Obtenção das amostras e extração de DNA genômico

171 Os animais foram coletados, mantidos, e posteriormente analisados de acordo com a
 172 legislação de proteção ambiental brasileira (permissão para coleta MMA/IBAMA/SISBIO –
 173 número 3245), com o Colégio Brasileiro de Experimentação Animal (COBEA), e com a
 174 aprovação do Instituto de Biociências (protocolos 405 e 504).

175 Para a obtenção das bibliotecas genômicas das espécies, extraímos o DNA a partir de
 176 tecido muscular, fígado ou sangue, previamente preservados em etanol 100%, utilizando os
 177 kits *Wizard Genomic DNA Purification Kit* (da Promega) e *DNeasy Blood & Tissue* (da
 178 QIAGEN), seguindo as instruções dos fabricantes. Além disso, realizamos uma etapa
 179 adicional para a remoção de RNA, com RNase A (da Invitrogen).

180 Os produtos da extração foram observados em gel de eletroforese composto por
 181 agarose na concentração de 1%, corado com *SYBR® Safe DNA* (0,015µL/mL), visualizados
 182 com a aplicação de luz ultravioleta (UV) de um transluminador, e fotografados com câmera
 183 Olympus C-5060, o que nos permitiu verificar a integridade do material extraído. O marcador

Low DNA Mass Ladder, da Invitrogen, com peso e concentração moleculares padrões, foi utilizado na eletroforese, permitindo também inferir a qualidade do DNA extraído.

3.2 Análises bioinformáticas

Para as análises bioinformáticas utilizamos bibliotecas genômicas de *reads* curtos pareados (Tabela 2). A maior parte das análises bioinformáticas e de processamento de dados foi realizada em um servidor remoto localizado no Instituto de Biociências de Botucatu, no Laboratório de Biologia e Genética de Peixes do Departamento de Biologia Estrutural e Funcional da UNESP, sob os cuidados do professor Dr. Claudio de Oliveira. Outras análises computacionais incluem a identificação e caracterização dos satelitomas das espécies utilizando a ferramenta TAREAN (*Tandem Repeat Analyzer*), do portal Galaxy. Portanto, alguns recursos computacionais foram providos pelo projeto ELIXIR-CZ (LM2015047), parte da infraestrutura internacional ELIXIR, uma organização intergovernamental que reúne recursos para o estudo das ciências biológicas, como bancos de dados e armazenamento em nuvem, *softwares* e ferramentas de análise, e servidores de alta capacidade computacional.

3.2.1 Isolamento de DNAs satélites com o TAREAN

Inicialmente, as bibliotecas foram preparadas, removendo-se os adaptadores, fazendo uma seleção da qualidade dos *reads* sequenciados, e padronizando-os para um tamanho de 100 pares de base.

A identificação dos DNAs satélites presentes no genoma das espécies foi realizada utilizando o TAREAN (NOVÁK et al., 2017), uma ferramenta do RepeatExplorer que automatiza o processo de análise dos gráficos obtidos das sequências repetitivas. Foram executadas diversas iterações desta ferramenta com as configurações padrões da mesma, utilizando 500.000 *paired-end reads*. Ao final de cada iteração, o programa produz e retorna uma lista com os prováveis DNAs satélites encontrados, com alta e baixa confiança. As sequências consenso destes satélites encontrados (que são as sequências com o maior pontuação de cobertura de *k-mers*) foram extraídas e inseridas em uma biblioteca customizada, para fazer a filtragem da biblioteca genômica aplicada no TAREAN anteriormente àquela rodada, a partir de um script customizado “*deconseq_run.py*” (SCHMIEDER; EDWARDS, 2011), a fim de remover os reads semelhantes aos satélites já encontrados. Assim, esta biblioteca filtrada era utilizada na próxima rodada do TAREAN em busca de novos satélites. Desta forma cada espécie teve um variado número de iterações da ferramenta até que nenhum novo DNA satélite fosse encontrado. Um esquema simplificado

do funcionamento do TAREAN encontra-se a seguir, na Figura 2 (adaptado de NOVÁK et al., 2017).

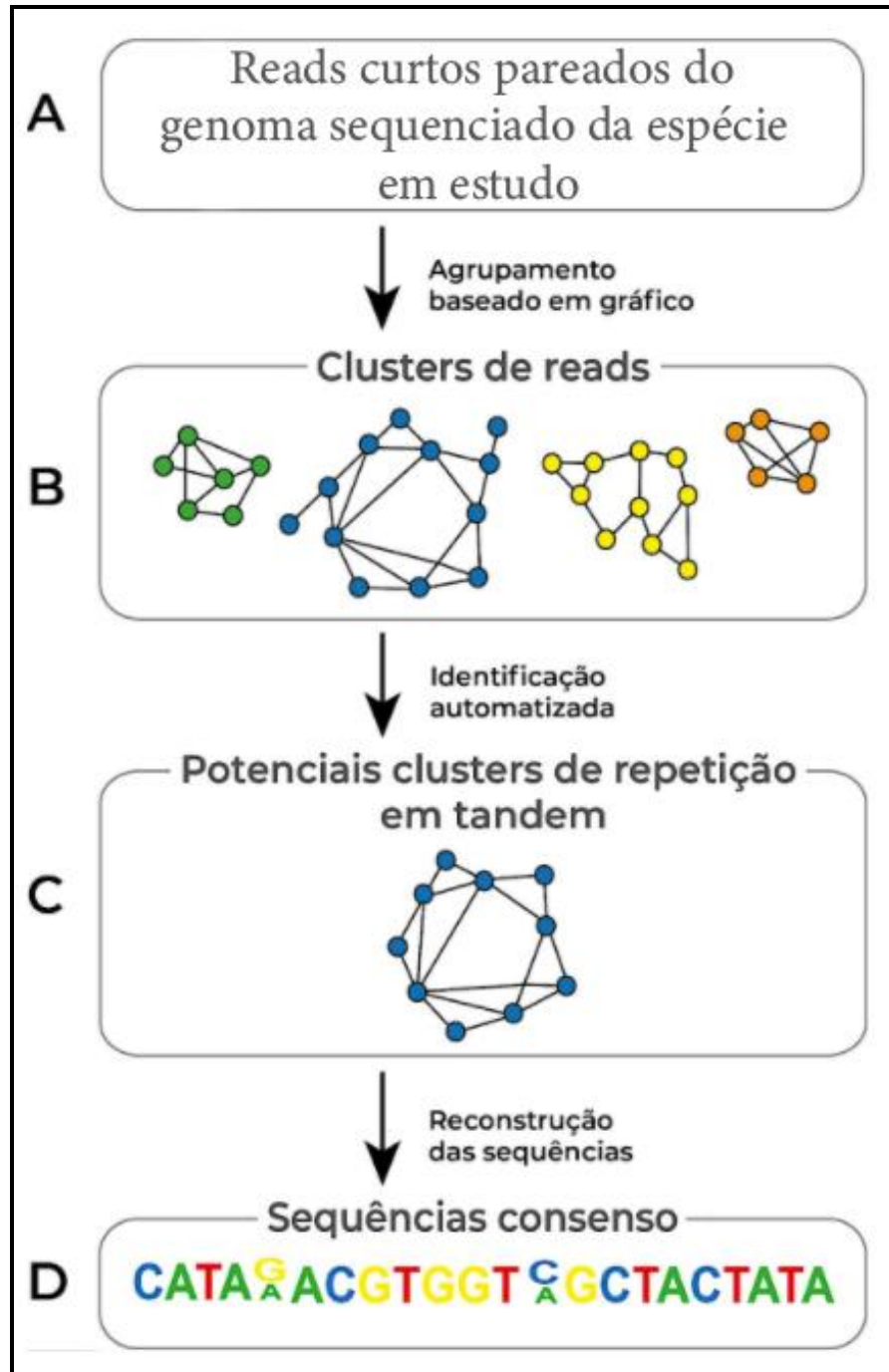


Figura 2: Workflow simplificado da ferramenta TAREAN (adaptado de NOVÁK et al., 2017). A) Inserção de *reads* curtos pareados de sequenciamento de nova geração; B) Agrupamento dos *reads* baseado em similaridades e sobreposição (90% de similaridade e pelo menos 55% do tamanho total dos *reads*); C) Identificação automatizada de gráficos que potencialmente representam satélites em tandem (estruturas circulares com os *reads*

densamente conectados); D) Reconstrução da sequência consenso das famílias de DNAs satélites a partir da decomposição em *k-mers* e de estimativas de pontuação (*score* de cobertura de *k-mers*) dos gráficos identificados na etapa anterior.

3.2.2 Caracterização do satelitoma das espécies

Após a caracterização de todos os possíveis DNAs satélites de uma determinada espécie, uma etapa de filtragem manual foi realizada para a remoção de outras sequências repetidas em tandem, como os elementos das famílias multigênicas do DNA ribossomal 45S, genes U de pequenos RNAs nucleares, e genes para histonas.

Em seguida, medimos as similaridades entre os DNAs satélites por meios de algoritmos de alinhamento e montagem, usando a ferramenta global de alinhamento do *software* Geneious (KEARSE et al., 2012), versão 8.1, e o script personalizado “rm_homology.py” (<https://github.com/fjruiaruano/ngs-protocols>), descrito em Ruiz-Ruano et al. (2016), para agrupá-los como satélites de uma superfamília (quando a similaridade é maior ou igual a 50%, indicando uma possível origem a partir de uma mesma sequência ancestral) ou como variantes (quando a similaridade é maior ou igual a 80%), critérios estabelecidos em Ruiz-Ruano et al. (2016). Nos casos de sequências extremamente similares (i.e. similaridade maior ou igual a 95%), as mesmas foram consideradas sendo a mesma variante, e portanto, uma delas (a de menor abundância) era removida para evitar redundâncias no catálogo final.

Para estimar a abundância relativa de cada satélite no genoma das espécies (o número de *reads* mapeados dividido pelo número de nucleotídeos analisados) e sua divergência média, foi utilizado o *software* RepeatMasker (SMIT et al., 2017), que utiliza uma implementação do algoritmo Smith-Waterman-Gotoh para comparar as sequências nucleotídicas inseridas, no caso, em subamostras das bibliotecas das espécies contendo 5 milhões de *reads* curtos pareados, exceto em *C. gomesi* e em *L. friderici*, onde o tamanho das bibliotecas foram limitantes, e utilizamos subamostras de 4 milhões e de 1 milhão de *reads* curtos pareados, respectivamente.

Os satélites foram nomeados de acordo com a sua abundância no genoma, com os critérios estabelecidos em Ruiz-Ruano et al. (2016), sendo o primeiro satélite do catálogo aquele mais abundante, e assim por diante. Por exemplo, o satélite AmeSat01-1633 é o satélite mais abundante caracterizado em *A. mexicanus*, e sua unidade de repetição possui 1633 pb de comprimento. Excepcionalmente, em *C. gomesi* os novos satélites encontrados pelo TAREAN foram adicionados na sequência dos satélites anteriormente descritos em Serrano-Freitas et al. (2020).

A correlação estatística (coeficiente de correlação de Pearson) entre os traços dos satelitomas foi estimada por meio de um script próprio customizado, construído na linguagem Python 3, que utiliza as bibliotecas seaborn (WASKOM, 2021) e matplotlib (HUNTER, 2007) para realizar as medidas e fazer a plotagem do gráfico do tipo *heatmap* para visualização.

3.2.3 Comparações entre os catálogos das espécies

Considerando que diferentes análises filogenômicas evidenciaram que Crenuchidae é grupo-irmão de todas as outras famílias de Characoidei (OLIVEIRA et al., 2011; ARCILA et al., 2017; MELO et al., 2022), partimos do princípio de que, caso algum satélite estivesse compartilhado entre *C. gomesi* e qualquer outra espécie em análise, seria potencialmente um satélite conservado.

Antes de comparar os catálogos de satDNAs entre espécies, nós realizamos uma etapa de clusterização manual em *C. gomesi*, uma vez que uma certa abundância genômica é necessária para que uma sequência seja caracterizada pela pipeline do TAREAN. Nesse sentido, realizamos uma busca na biblioteca de *C. gomesi* por todos os satélites descritos para as outras espécies do estudo (totalizando 841 satDNAs), utilizando o *software* RepeatMasker, identificando e isolando todos os *reads* da biblioteca genômica que apresentaram similaridade com estes satélites, e gerando uma subamostra destes *reads* da biblioteca genômica com um *script* personalizado e gentilmente cedido pelo Dr. Francisco J. Ruiz-Ruano. Com esta subamostra, utilizamos o TAREAN em busca de caracterizar satélites adicionais, que não foram identificados na análise anterior, mas que se encontram presentes na biblioteca de *C. gomesi*, para serem utilizados nas análises comparativas dos satelitomas. Novamente, estes novos satélites foram adicionados ao final do catálogo.

Com os catálogos de todas as espécies de estudo já caracterizados, foi possível construir uma biblioteca com todos os satélites para realizar a comparação total, a fim de verificar a existência de algum caso adicional de um DNA satélite compartilhado por *C. gomesi* e diferentes espécies da subordem dos Characoidei. Para isso foram utilizados o script “rm_homology.py”, resultando em todas as similaridades existentes entre os satélites, e a ferramenta de montagem *de novo* implementada no *software* Geneious 8.1, resultando em *contigs* de satélites que apresentam similaridades, nos indicando uma possível relação. Todos os dados obtidos de possíveis relações entre os satélites também foram visualmente inspecionados com o auxílio da ferramenta de alinhamento do Geneious 8.1.

Assim, as sequências descobertas conservadas foram selecionadas para uma análise do perfil do satélite em todas as espécies do estudo utilizando o programa RepeatProfiler (NEGM et al., 2021). Este programa mapeia os *reads* das bibliotecas selecionadas na referência desejada (no caso, a sequência dos satélites repetida até passar do tamanho de 200 pares de base) utilizando a ferramenta de alinhamento Bowtie2 (LANGMEAD; SALZBERG, 2012), além do conjunto de ferramentas SAMtools (LI et al., 2009) e das linguagens de programação Python 3 (VAN ROSSUM; DRAKE, 2009) e R (TEAM, 2013) para processar e permitir a visualização dos dados obtidos. Uma ressalva importante é a utilização do parâmetro “–singlecopy” no RepeatProfiler, o que permite a utilização conjunta de genes de cópia única na referência, nos retornando um número de cópias mapeadas sobre a referência já normalizado. Os genes utilizados e seus respectivos números de acesso são: *coq6* [XM_017711829.1], *foxl2* [XM_007232295.3], *gastrula zinc finger* [XM_022685636.1], *lactamase* [XM_022682177.1], *msh4* [XM_017711771.1], *ppfia1* [XM_022685633.1], *prospero* [XM_017708821.1], *tubulin-kinase* [XM_017711762.1], *zdhhc22* [XM_017711775.1], e *znf106* [XM_017711848.1].

Para esta análise também foram utilizadas subamostras das bibliotecas de cada espécie com 5 milhões de *reads* curtos pareados, exceto nos casos onde o tamanho da biblioteca era um fator limitante, como em *C. gomesi* e *P. nattereri*, ambas subamostras com 4 milhões de *reads* curtos pareados, em *H. malabaricus*, com 3,8 milhões de *reads* curtos pareados, além de *M. asterias*, *P. mesopotamicus*, *H. gracilis* e *L. friderici*, com subamostras de 3 milhões, 2 milhões, 1,9 milhão e 1 milhão de *reads* curtos pareados, respectivamente.

4. RESULTADOS

Descrevemos, com o uso da ferramenta TAREAN, 5 novos catálogos de DNAs satélites das espécies *H. malabaricus*, *L. friderici*, *C. modestus*, *P. semifasciata* e *A. mexicanus*, todas distribuídas em diferentes famílias da subordem Characoidei. Adicionalmente, complementamos o catálogo de *C. gomesi*, anteriormente descrito utilizando a ferramenta RepeatExplorer1 e o satMiner, com novas famílias de satélites. Por fim, realizamos uma análise comparativa destes novos catálogos em conjunto com outros anteriormente publicados, totalizando 15 espécies distintas.

4.1 Descrição de DNAs satélites nas diferentes espécies

Para a complementação do catálogo de *C. gomesi* com novos satélites isolados a partir do TAREAN, foram realizadas 8 iterações do *workflow* envolvendo a ferramenta e o deconseq, até que nenhum novo satélite em potencial fosse encontrado. Foram detectados 58 novos satélites, além dos 59 descritos anteriormente em Serrano-Freitas et al. (2020). Na etapa de remoção de outros tipos de elementos repetitivos, verificamos a ocorrência de uma cópia perfeita do DNAr 5S no centro do satélite CgoSat09-236, e por conta disso, o satélite foi desconsiderado nas análises posteriores. A caracterização adicional de satélites envolvendo o RepeatMasker e o TAREAN, usando como “iscas” os satDNAs presentes em outras espécies, resultou em 23 satélites adicionais em *C. gomesi*. Assim, o catálogo final desta espécie foi composto por 139 satélites válidos, representando 8,1% do genoma total (Tabela 3). É o catálogo descrito no estudo com o maior número de satélites, apresentando 15 superfamílias. A maioria (71) dos satélites são grandes (maiores que 100 pares de base), com média e mediana do RUL (*Repeat unit length* - tamanho da unidade de repetição da sequência) de 341 e 103 pares de base, respectivamente, e uma mediana do conteúdo A+T de 63%, indicando uma prevalência de satélites ricos em A e T.

Tabela 3: As características principais dos satélites de *C. gomesi*. SF – superfamília; RUL – o comprimento da unidade de repetição; A+T – o conteúdo relativo de adenina e timina.

SF	Satélite	RUL	Divergência	Abundância	A+T(%)
	CgoSat01-184	184	11,71	0,0042342	62
3	CgoSat02-52	52	7,53	0,0030719	71
	CgoSat03-19	19	6,42	0,0020133	68
1	CgoSat04-236	236	11,51	0,0019645	64
	CgoSat05-43	43	2,11	0,0014637	72
	CgoSat06-60	60	6,71	0,0013877	53
4	CgoSat07-67	67	4,58	0,0011776	64
1	CgoSat08-236	236	12,73	0,0011029	66
	CgoSat10-42	42	13,53	0,0010053	52
	CgoSat11-171	171	12,84	0,000935	70
	CgoSat12-6-tel	6	15,62	0,0008692	50
	CgoSat13-61	61	4,41	0,0008053	54
	CgoSat14-46	46	4,61	0,0007919	74
	CgoSat15-176	176	10,44	0,0007688	65
	CgoSat16-43	43	21,22	0,0007602	56
	CgoSat17-39	39	2,59	0,0006996	77
	CgoSat18-25	25	3,82	0,0006548	72
2	CgoSat19-237	237	10,48	0,0006432	66
	CgoSat20-62	62	3,15	0,00063	81

	CgoSat21-117	117	19,6	0,0006225	61
5	CgoSat22-178	178	5,66	0,0006106	61
1	CgoSat23-247	247	3,21	0,0004909	65
1	CgoSat24-230	230	8,74	0,0004853	61
	CgoSat25-281	281	9,43	0,000485	64
	CgoSat26-29	29	15,27	0,0004815	48
	CgoSat27-32	32	1,88	0,0003915	59
1	CgoSat28-245	245	7,45	0,000384	66
1	CgoSat29-245	245	5,16	0,0003767	66
	CgoSat30-34	34	4,09	0,0003501	71
	CgoSat31-103	103	12,02	0,0003271	55
6	CgoSat32-21	21	14,96	0,0003195	76
	CgoSat33-84	84	19,82	0,0003167	55
	CgoSat34-43	43	6,31	0,0003115	51
2	CgoSat35-677	677	4,3	0,0003087	71
	CgoSat36-114	114	6,28	0,0002566	47
1	CgoSat37-247	247	7,44	0,0002539	68
	CgoSat38-71	71	3,74	0,0002537	76
5	CgoSat39-177	177	6,13	0,000247	64
	CgoSat40-21	21	10,33	0,0002399	62
	CgoSat41-146	146	10,34	0,0002392	69
	CgoSat42-67	67	7,15	0,0002291	67
	CgoSat43-48	48	3,34	0,0002238	79
4	CgoSat44-177	177	12,36	0,0002175	72
1	CgoSat45-437	437	5,9	0,0002158	67
	CgoSat46-48	48	4,64	0,0002126	85
	CgoSat47-115	115	5,21	0,0002016	44
	CgoSat48-76	76	0,65	0,0001528	80
	CgoSat49-57	57	4,99	0,0001514	70
2	CgoSat50-227	227	7,05	0,0001431	66
	CgoSat51-32	32	2,37	0,00014	72
	CgoSat52-143	143	9,35	0,0001356	73
1	CgoSat53-236	236	17,86	0,0001287	65
	CgoSat54-103	103	3,86	0,0001238	37
	CgoSat55-47	47	7,52	0,0001087	47
7	CgoSat56-109	109	4,97	0,0000946	39
	CgoSat57-153	153	9,77	0,0000896	44
	CgoSat58-43	43	5,59	0,0000414	54
	CgoSat59-103	103	12,94	0,0000047	64
	CgoSat60-2063	2063	7,81	0,009519278	61
8	CgoSat61-1957	1957	3,63	0,003223369	61
	CgoSat62-563	563	6,1	0,00313754	55
8	CgoSat63-1972	1972	5,49	0,002832876	63

9	CgoSat64-22	22	19,72	0,002206022	50
	CgoSat65-5621	5621	6,65	0,002119854	64
9	CgoSat66-21	21	20,13	0,001779052	48
	CgoSat67-1094	1094	14,28	0,001338686	63
10	CgoSat68-767	767	13,25	0,001050766	66
9	CgoSat69-53	53	5,15	0,000812277	55
	CgoSat70-760	760	7,04	0,000695999	50
6	CgoSat71-21	21	17,68	0,000630494	76
11	CgoSat72-1521	1521	3,58	0,000623512	55
	CgoSat73-1319	1319	3,4	0,000622291	57
12	CgoSat74-1471	1471	0,98	0,000537865	58
	CgoSat75-1560	1560	1,06	0,000495083	60
13	CgoSat76-865	865	1,67	0,000385101	64
13	CgoSat77-613	613	3,03	0,000359584	57
11	CgoSat78-1169	1169	2,88	0,000352204	57
	CgoSat79-2223	2223	1,67	0,00034288	38
3	CgoSat80-57	57	4,96	0,000339984	67
	CgoSat81-773	773	2,42	0,00029879	66
	CgoSat82-1130	1130	3,23	0,000291599	59
14	CgoSat83-42	42	6,77	0,000281572	67
	CgoSat84-23	23	9,29	0,000219739	61
	CgoSat85-22	22	10,73	0,000210421	64
15	CgoSat86-607	607	11,37	0,000204157	58
13	CgoSat87-818	818	2,46	0,00019876	59
	CgoSat88-273	273	7,72	0,000186397	64
12	CgoSat89-1175	1175	8,14	0,000182007	60
16	CgoSat90-642	642	6,68	0,000181882	61
	CgoSat91-110	110	13,02	0,000181004	57
10	CgoSat92-624	624	1,49	0,000172311	60
	CgoSat93-495	495	1,99	0,000168965	65
7	CgoSat94-224	224	2,81	0,000159889	42
	CgoSat95-438	438	6,17	0,000155608	57
	CgoSat96-75	75	15,37	0,000148938	63
16	CgoSat97-675	675	0,98	0,000147327	57
	CgoSat98-111	111	11,2	0,000135557	54
14	CgoSat99-15	15	11,17	0,000133444	67
	CgoSat100-64	64	2,88	0,000125537	66
	CgoSat101-146	146	14,46	0,000124726	58
2	CgoSat102-237	237	4,89	0,000119127	68
	CgoSat103-31	31	10,79	0,000115797	71
	CgoSat104-327	327	4,77	0,000112819	59
	CgoSat105-183	183	9,55	0,000110567	59
14	CgoSat106-42	42	6,34	0,000109308	62

	CgoSat107-25	25	4,82	0,000108252	68
	CgoSat108-43	43	11,86	0,000106274	70
	CgoSat109-33	33	16,74	9,65817E-05	70
	CgoSat110-42	42	22,15	9,02748E-05	48
	CgoSat111-25	25	9,14	8,84443E-05	68
	CgoSat112-57	57	3,24	8,82252E-05	53
	CgoSat113-33	33	5,88	8,52277E-05	52
	CgoSat114-16	16	5,84	8,36782E-05	69
	CgoSat115-31	31	8,23	8,0875E-05	71
	CgoSat116-33	33	9,87	7,89814E-05	64
	CgoSat117-60	60	1,54	7,51646E-05	80
1	CgoSat118-165	165	12,01	0,00293196	62
	CgoSat119-705	705	9,00	0,00146433	65
	CgoSat120-657	657	11,55	0,000664552	57
	CgoSat121-44	44	10,50	0,000414043	61
15	CgoSat122-1208	1208	9,02	0,000308881	57
	CgoSat123-756	756	1,49	0,000246487	67
	CgoSat124-179	179	14,03	0,000201402	54
	CgoSat125-36	36	26,68	0,00015792	61
	CgoSat126-27	27	9,71	0,00013528	59
14	CgoSat127-21	21	17,84	0,000102094	71
	CgoSat128-20	20	17,12	8,22067E-05	65
	CgoSat129-32	32	5,91	7,46262E-05	68
	CgoSat130-50	50	10,15	6,97512E-05	70
	CgoSat131-62	62	5,34	6,90816E-05	74
	CgoSat132-96	96	10,15	5,86596E-05	59
	CgoSat133-173	173	7,16	4,04307E-05	54
	CgoSat134-15	15	17,47	3,78156E-05	53
	CgoSat135-21	21	7,28	3,62772E-05	61
	CgoSat136-30	30	15,45	3,39728E-05	63
	CgoSat137-130	130	12,82	3,30668E-05	71
	CgoSat138-62	62	6,53	3,3021E-05	66
	CgoSat139-48	48	15,31	2,27896E-05	62
	CgoSat140-94	94	8,19	8,67822E-06	57

345

346

347

348

349

350

351

As espécies *H. malabaricus*, *L. friderici*, *C. modestus*, *P. semifasciata* e *A. mexicanus* tiveram seus satelitomas descritos pela primeira vez. Em *H. malabaricus* (Tabela Suplementar 1) realizamos 5 iterações do *workflow* que envolve o TAREAN e o deconseq, o menor número de iterações realizado para uma espécie neste estudo. Para *L. friderici* (Tabela Suplementar 2) e *P. semifasciata* (Tabela Suplementar 3) foram realizadas 7 iterações. E para *C. modestus* (Tabela Suplementar 4) e *A. mexicanus* (Tabela Suplementar 5) foram realizadas

9 iterações do *workflow*. As principais métricas de todos os satelitomas utilizados no estudo, inclusive aqueles descritos anteriormente em outros trabalhos, estão sumarizadas na Tabela 4.

Tabela 4: Estatísticas dos satelitomas utilizados no estudo. Satélites grandes – satélites maiores que 100 pb; Div. média – média da divergência de Kimura dos satélites; % genoma – porção do genoma total representada pelos satélites do catálogo.

Espécies	Nº de satélites	Satélites grandes	RUL médio (pb)	Mediana RUL (pb)	Div. média	% genoma	Mediana A+T (%)
<i>C. gomesi</i>	139	71	341	103	8,42	8,1	63
<i>H. malabaricus</i>	56	39	493	228	6,09	10,3	58
<i>H. gracilis</i>	59	29	281	94	7,82	4,3	57
<i>P. nattereri</i>	57	29	472	115	7,92	3	57
<i>M. asterias</i>	31	9	113	62	10,2	3,3	60
<i>C. macropomum</i>	46	27	429	173,5	8,77	6,2	60
<i>P. brachypomus</i>	63	47	862	515	9,02	7,7	59
<i>P. mesopotamicus</i>	30	14	286	87,5	10,06	4,2	60
<i>L. friderici</i>	74	29	210	53,5	8,26	8,5	59,35
<i>P. lineatus</i>	51	20	349	59	8,25	5,1	54,8
<i>C. modestus</i>	105	40	233	52	8,29	8,2	55
<i>P. semifasciata</i>	71	44	391	182	7,43	6	61,4
<i>T. auritus</i>	53	10	121	61	7	6,1	59,4
<i>P. paranae</i>	101	39	286	75	8,39	12,6	57,1
<i>A. mexicanus</i>	45	28	327	175	8,24	3,8	59,4

Após a complementação do satelitoma, a espécie *C. gomesi*, com 139 famílias de DNAs satélites, possui o maior catálogo entre todos os satelitomas analisados neste estudo, embora apresente menos famílias do que *M. macrocephalus*, com 164 famílias (UTSUNOMIA et al., 2019), e *M. elongatus*, com 140 famílias (CREPALDI; PARISE-MALTEMPI, 2021), ambos caracterizados com o *workflow* que envolve o RepeatExplorer1 e o satMiner. A porcentagem do genoma composta pelos satélites variou de 3%, em *P. nattereri* a 12,6% em *P. paranae*, que teve uma representatividade genômica de satélites menor apenas do que a biblioteca genômica de fêmea de *M. macrocephalus* (UTSUNOMIA et al., 2019), com 13,4%. Os dados obtidos não apresentam nenhum padrão aparente de relação entre número de satélites ou porcentagem do genoma composta por satélites e o fato das espécies serem mais basais ou mais derivadas.

Das 15 espécies do estudo, sete possuem uma maioria de satélites grandes (RUL maior que 100 bp), demonstrando praticamente um equilíbrio entre satélites pequenos e grandes num geral, mesmo que a mediana do RUL de toda a amostra de 981 satélites seja de 91 pares

de base. O satelitoma de *P. brachypomus* apresentou uma mediana bem alta comparado aos outros catálogos, apresentando vários satélites bem maiores que ultrapassam 1000 pares de base. Quanto ao conteúdo A+T, todos os satelitomas apresentaram uma mediana maior que 54,8%, indicando uma predominância de satélites mais ricos em A e T.

Nas análises de correlação estatística dos traços dos satelitomas, observamos a existência de poucos casos de correlação significativa entre as variáveis. RUL e divergência apresentaram uma correlação moderada (maior que 0.35 – TAYLOR, 1990) para *A. mexicanus* (coeficiente de correlação de Pearson igual a -0,36), *H. gracilis* (-0,53), *M. asterias* (-0,5) e *P. lineatus* (-0,37). Os traços RUL e conteúdo A+T apresentaram uma correlação moderada apenas para *H. gracilis* (-0,41). RUL e abundância apresentaram uma correlação moderada apenas para *C. gomesi* (0,36). Em todos os outros casos, em termos de espécies e de traços dos satelitomas (divergência e abundância, divergência e conteúdo A+T, e abundância e conteúdo A+T), foram observadas apenas correlações fracas. Todos os coeficientes podem ser observados na Tabela Suplementar 6. Quando unimos todos os catálogos das 15 espécies do estudo, nenhuma correlação estatística significativa foi observada (Figura 3). Portanto, nossos dados não indicam a existência de um padrão de correlação linear entre os traços dos DNAs satélites neste grupo de peixes.

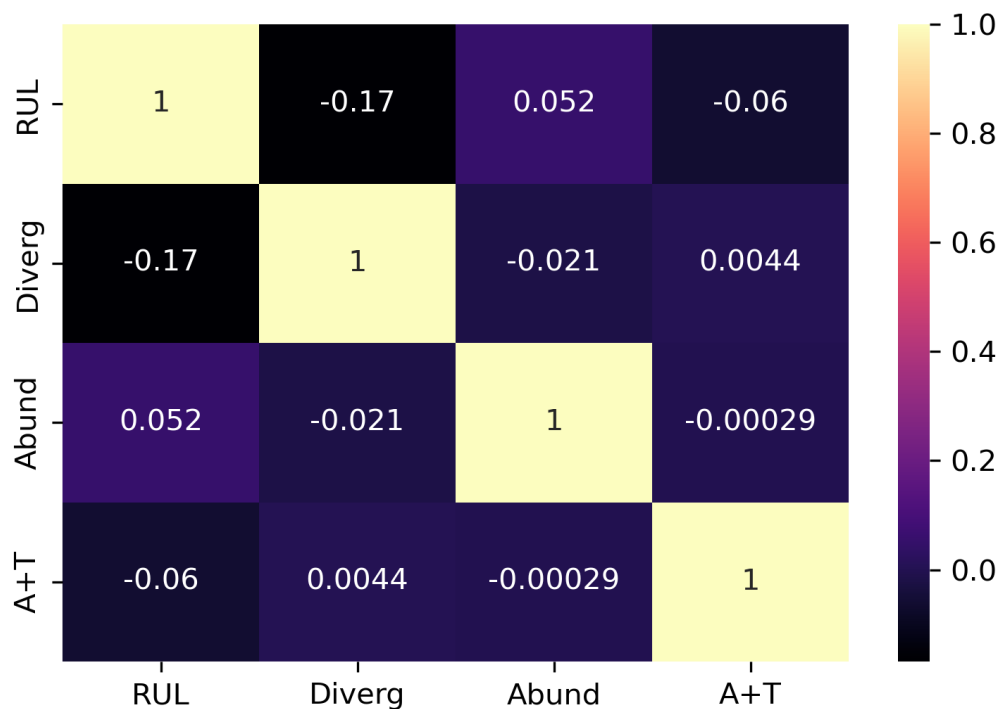


Figura 3: Gráfico do tipo *heatmap* demonstrando os coeficientes de correlação de Pearson entre os traços dos satelitomas de todas as 15 espécies de Characoidei analisadas.

4.2 Análises comparativas dos catálogos de satélites

As análises de similaridade revelaram que, de um catálogo final de 981 satDNAs, cinco famílias de satDNAs de *C. gomesi* apresentaram identidade elevada (maior que 80%) e, consequentemente, uma extrema conservação na composição, em relação a famílias de satélites descritas em outras espécies, sendo estas CgoSat71-21, CgoSat26-29, CgoSat10-42, CgoSat66-21, e CgoSat32-21. Além disso, destacam-se dois casos adicionais com alta similaridade entre monômeros de satélites de *C. gomesi* e de outras espécies, porém com valores de identidade inferiores a 80%, sendo estes CgoSat06-60 e CgoSat130-50.

As análises dos perfis desses satélites de *C. gomesi* compartilhados com outras espécies, em todas as bibliotecas genômicas analisadas, utilizando o RepeatProfiler, revelou a ocorrência de *reads* mapeados em quase todas as espécies, inclusive nas espécies em que determinado satélite não estava caracterizado em seus respectivos catálogos *a priori*. Este resultado revelou que os satélites conservados analisados estão presentes em todas as superfamílias de Characoidei das quais possuímos amostras. O número máximo de cópias mapeadas dos 7 satélites de *C. gomesi* em todas as espécies, e o número médio de cópias mapeadas por toda a referência, estão sumarizados nas tabelas 5 e 6, respectivamente.

Tabela 5: Pico máximo de cópias mapeadas dos satélites selecionados nas bibliotecas das espécies estudadas.

	Pico do número de cópias mapeadas de cada satélite						
Espécie	CgoSat26-29	CgoSat10-42	CgoSat71-21	CgoSat66-21	CgoSat32-21	CgoSat06-60	CgoSat130-50
<i>C. gomesi</i>	5.139	3.022	1.983	7.935	2.308	20.609	209
<i>H. malabaricus</i>	27	102	27	15	111	1.048	22
<i>H. gracilis</i>	624	434	1.386	129	185	214	321
<i>P. nattereri</i>	1.106	194	1.494	202	106	1.234	265
<i>M. asterias</i>	1.193	227	1.158	191	75	786	302
<i>C. macropomum</i>	2.444	313	1.761	447	138	2.516	261
<i>P. brachypomus</i>	2.786	238	1.672	506	115	1.706	265
<i>P. mesopotamicus</i>	2.279	590	5.036	1.003	459	2.147	550
<i>L. friderici</i>	776	505	603	40	132	285	256
<i>P. lineatus</i>	530	221	981	346	32	492	12

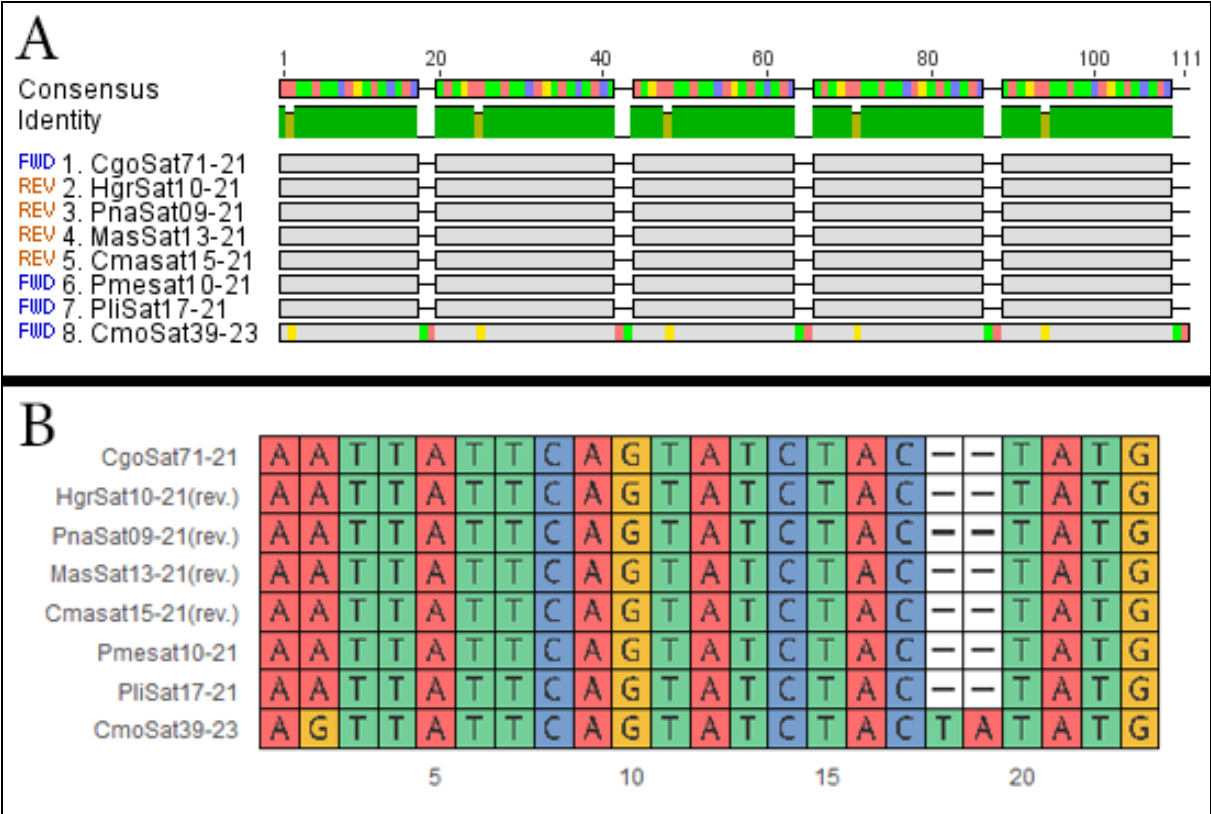
<i>C. modestus</i>	506	160	172	271	709	155	41
<i>P. semifasciata</i>	66	89	7	245	33	28.503	0
<i>T. auritus</i>	2.231	338	1.952	331	1.436	2.494	2
<i>P. paranae</i>	1.220	76	416	101	1.495	9	26
<i>A. mexicanus</i>	791	84	397	38	761	9	7

Tabela 6: Média do número de cópias mapeadas dos satélites selecionados nas bibliotecas das espécies estudadas.

Espécie	Média do número de cópias mapeadas de cada satélite						
	CgoSat26-29	CgoSat10-42	CgoSat71-21	CgoSat66-21	CgoSat32-21	CgoSat06-60	CgoSat130-50
<i>C. gomesi</i>	2119,96	1250,47	845,26	3106,40	990,84	9842,82	90,94
<i>H. malabaricus</i>	10,69	36,96	10,48	5,55	44,57	381,73	9,78
<i>H. gracilis</i>	256,98	174,50	587,44	46,77	78,52	86,31	124,52
<i>P. nattereri</i>	435,65	77,83	626,01	70,39	42,85	538,83	101,94
<i>M. asterias</i>	472,46	93,31	504,27	64,25	33,16	309,11	136,09
<i>C. macropomum</i>	961,03	123,76	783,62	169,14	59,44	1060,19	110,21
<i>P. brachypomus</i>	1048,60	94,04	712,82	180,20	51,23	728,47	116,49
<i>P. mesopotamicus</i>	978,31	238,73	2156,03	347,54	194,97	912,14	210,87
<i>L. friderici</i>	342,45	211,35	275,47	13,20	50,88	106,34	118,98
<i>P. lineatus</i>	221,71	90,03	434,42	133,56	13,38	212,02	4,25
<i>C. modestus</i>	218,43	58,60	72,01	100,19	308,90	59,65	17,11
<i>P. semifasciata</i>	23,53	33,62	1,90	104,62	13,88	10196,61	0
<i>T. auritus</i>	959,65	124,04	835,24	130,96	647,36	998,20	0,47
<i>P. paranae</i>	517,58	24,24	189,47	30,34	659,94	2,40	7,57
<i>A. mexicanus</i>	329,76	26,46	173,29	11,91	329,97	3,59	2,79

O primeiro caso (Alinhamento 1 - Figura 4) evidenciou a ocorrência de um DNA satélite praticamente idêntico, descrito independentemente em 8 espécies, o CgoSat71-21. A análise do perfil deste satélite apresentou cópias mapeadas em todas as espécies analisadas, confirmando a organização em tandem dos *arrays*. A biblioteca de *P. semifasciata* apresentou um número muito baixo de cópias, com um número médio de cópias mapeadas de 1,9, indicando uma diminuição dos *arrays* deste satélite no genoma da espécie pertencente à família Lebiasinidae, restando apenas algum resquício de sua existência. Os gráficos do perfil do satélite em algumas das bibliotecas genômicas estão dispostos na Figura 5, a fim de exemplificação.

427



428

429 **Figura 4:** Alinhamento 1: A) Alinhamento de 5 monômeros em tandem dos satélites que
430 apresentam alta similaridade com CgoSat71-21; B) As unidades de repetição de cada um dos
431 satélite alinhadas.

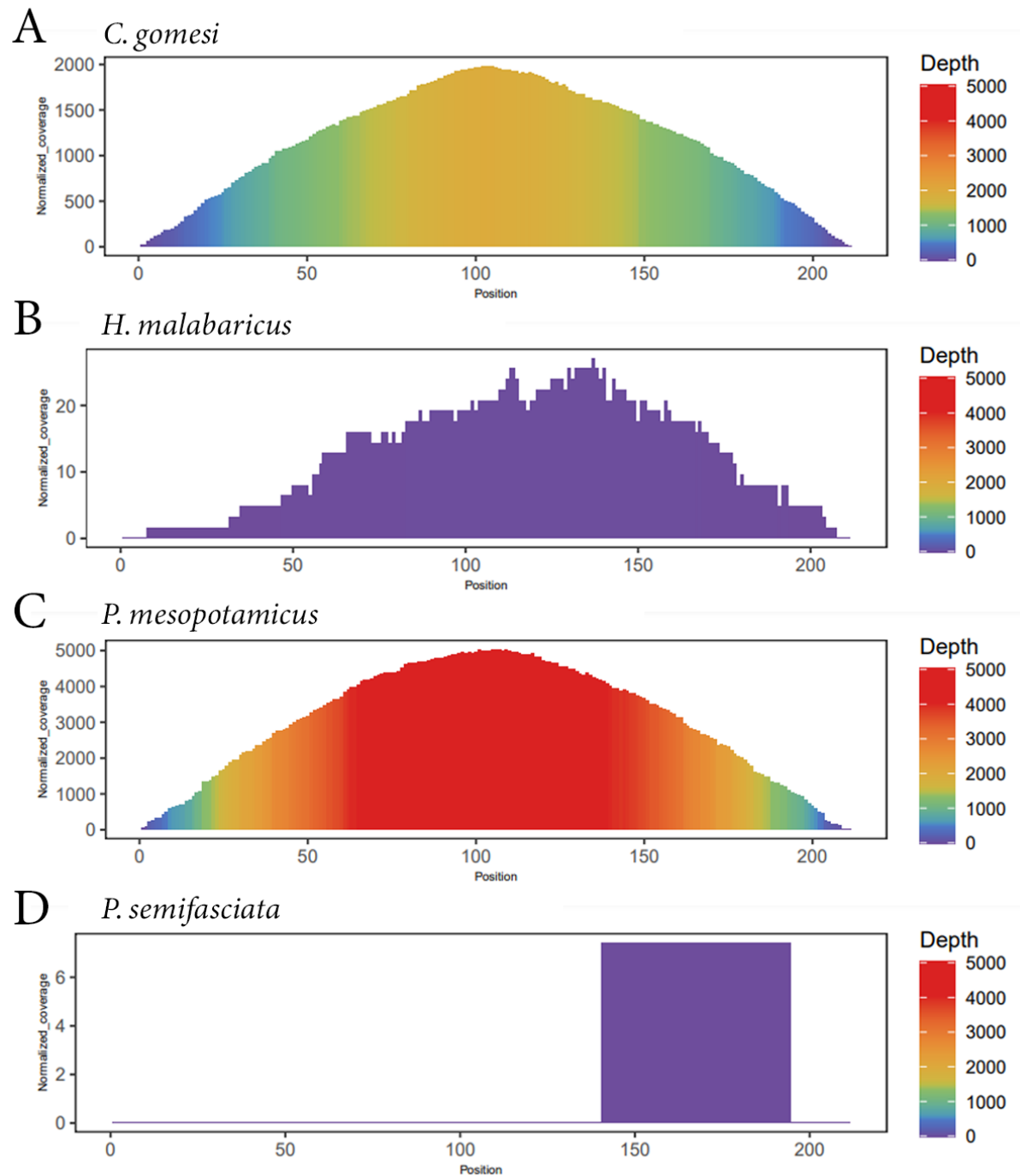


Figura 5: Gráficos de cobertura do CgoSat71-21 produzidos com o RepeatProfiler: A) Gráfico da biblioteca de *C. gomesi*, de onde o satélite foi isolado, com uma cobertura média/alta do satélite; B) Gráfico da biblioteca de *H. malabaricus*, com uma cobertura muito baixa do satélite, mas que ainda demonstra a organização dos arranjos em tandem do satélite; C) Gráfico da biblioteca de *P. mesopotamicus*, com uma cobertura alta do satélite, maior até do que em *C. gomesi*; D) Gráfico da biblioteca de *P. semifasciata*, com uma cobertura máxima de 7 cópias, porém com a cobertura média de 1,90, evidenciando a presença

resquicial do satélite, que já não apresenta uma organização em tandem na biblioteca genômica da espécie.

Os outros alinhamentos apresentam satélites descritos independentemente em alguns catálogos, e que possuem uma uma distribuição entre as espécies bastante similar ao caso do alinhamento 1. O alinhamento 2, do CgoSat26-29, apresenta satélites descritos em 10 diferentes espécies, variando de 27 a 31 pares de base (Figura 6). O satélite mais diferenciado no contig é o CmoSat32-31, que ainda assim apresenta uma identidade variando de 70 a 77% com os outros satélites, que por sua vez, apresentam uma identidade entre si que varia de 86 a 96%. A análise dos gráficos de perfil do satélite descrito em *C. gomesi* revelou um número de cópias mapeadas em todas as bibliotecas genômicas, demonstrando uma organização dos arrays em tandem nas espécies. Pode-se dizer que o satélite está presente, mesmo que em um baixo número de cópias, no genoma de todas as espécies de Characoidei analisadas.

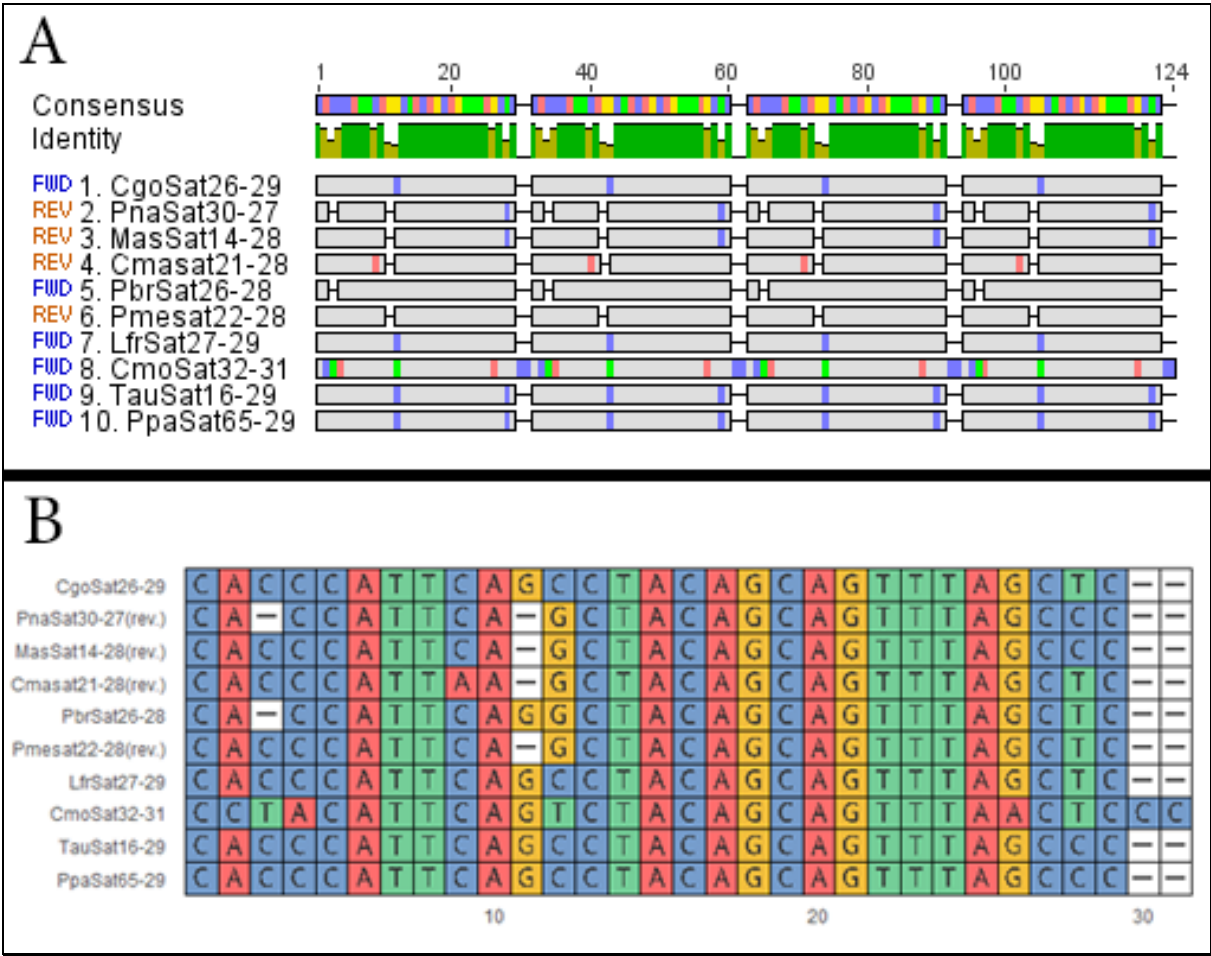


Figura 6: Alinhamento 2: A) Alinhamento de 4 monômeros em tandem dos satélites que apresentam alta similaridade com CgoSat26-29; B) As unidade de repetição de cada um dos satélite alinhadas.

O alinhamento 3 (Figura 7) inclui 11 satélites de 42 pares de base altamente similares descritos em diferentes espécies, dentre eles CgoSat10-42. A análise do perfil deste satélite mostrou um número considerável de cópias em todas as espécies analisadas, demonstrando uma organização dos *arrays* em tandem. Vale ressaltar que PmeSat07-42 foi mapeado nos cromossomos de *P. mesopotamicus* por meio de hibridação *in situ* fluorescente (FISH), no trabalho de Rodrigues (2021), onde demonstrou sinais, e portanto organização em *clusters*, nas regiões teloméricas de diversos cromossomos da espécie. PliSat12-42 também foi mapeado por FISH nos cromossomos de *P. lineatus* em Stornioli et al. (2021), mas não demonstrou nenhum sinal. Nossos dados são congruentes com estes resultados de mapeamento cromossômico anteriores, onde *P. mesopotamicus*, que demonstrou clusters do satélite, apresentou aqui um número maior de cópias deste satélite em comparação à *P. lineatus*, que não demonstrou *clusters* na análise com FISH.

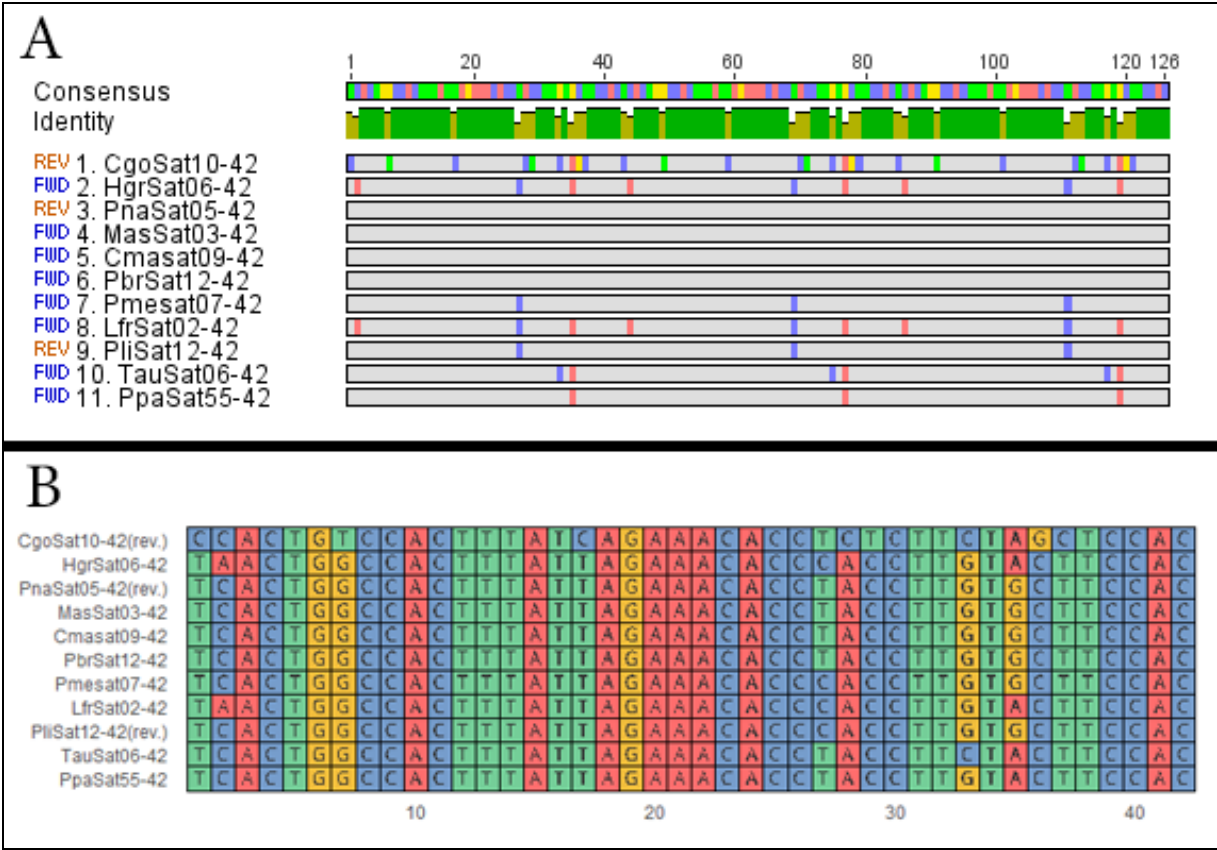


Figura 7: Alinhamento 3: A) Alinhamento de 3 monômeros em tandem dos satélites que apresentam alta similaridade com CgoSat10-42; B) As unidades de repetição de cada um dos satélite alinhadas.

O alinhamento 4 (Figura 8), do CgoSat66-21, apresentou satélites descritos em três espécies. CgoSat66-21 e PseSat45-21 apresentam uma identidade de 95%, porém a variante do satélite caracterizada em *C. macropomum* é bastante similar a dois monômeros da variante menor em tandem, com uma inserção adicional, apresentando uma distância de 81%. Esta organização da variante de 43 bp caracteriza uma HOR com subrepetições homólogas à variante menor, como no caso da super família 11 de satélites no gafanhoto *Pyrgomorpha conica* (RUIZ-RUANO et al., 2018), ou das diferentes HORs do *Alpha* Satélite em humanos (URALSKY et al., 2019). A análise do perfil do satélite de *C. gomesi* nas bibliotecas também revelou cópias mapeadas em praticamente todas as espécies, demonstrando um padrão de organização dos *arrays* em tandem, inclusive em *H. malabaricus*, com um número baixo de cópias mapeadas (média de 5,5), indicando uma diminuição dos *arrays* do satélite no genoma da referida espécie, da família Erythrinidae, mas este padrão de organização dos *arrays* ainda é observado.

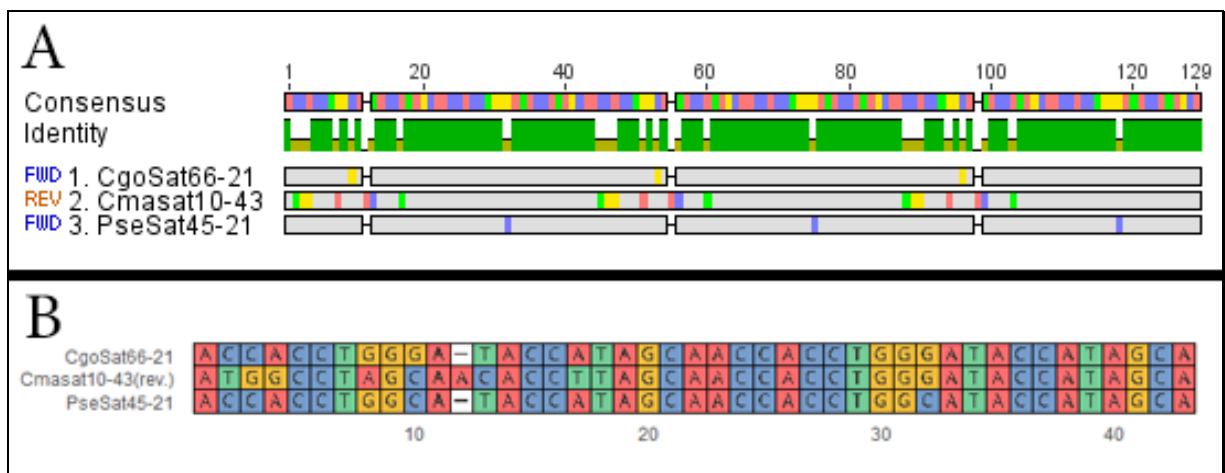


Figura 8: Alinhamento 4: A) Alinhamento de 6 monômeros em tandem dos satélites CgoSat66-21 e PseSat45-21, com 3 monômeros em tandem do CmaSat10-43; B) Uma unidade de repetição de CmaSat10-43 alinhada com um dímero de ambos os satélites menores.

O alinhamento 5 (Figura 9) inclui o CgoSat32-21, que também foi descrito independentemente em duas outras das espécies, *C. modestus* e *P. paranae*. O satélite de *C.*

gomesi teve reads mapeados nas bibliotecas genômicas de todas as espécies. Em *T. auritus*, sem o satélite em seu catálogo, o número de cópias mapeadas foi consideravelmente alto, inclusive, maior do que em *C. modestus*. O baixo número de cópias mapeadas em *P. lineatus* e em *P. semifasciata* indica uma diminuição acentuada do tamanho dos *arrays* especificamente nestas linhagens evolutivas.

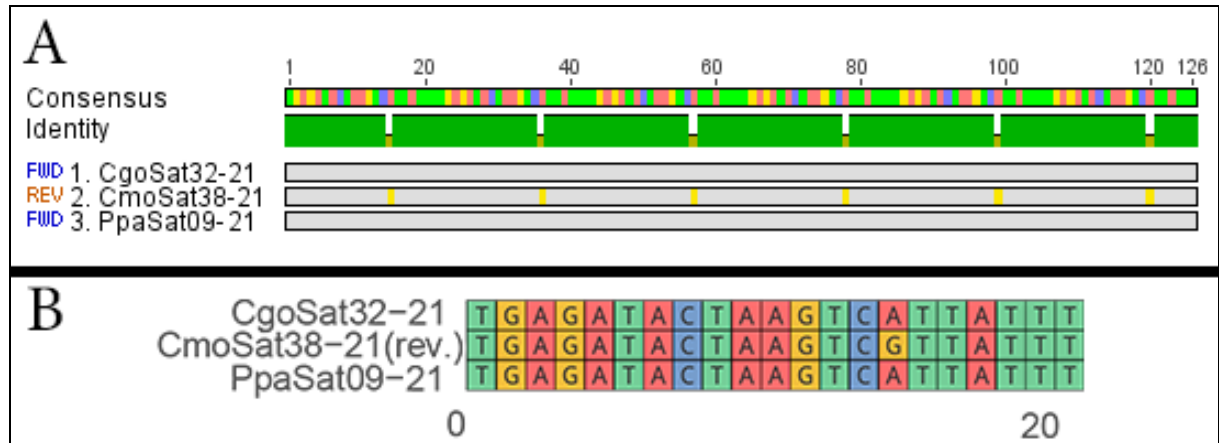


Figura 9: Alinhamento 5: A) Alinhamento de 6 monômeros em tandem dos satélites CgoSat32-21, CmoSat38-21 e PpaSat09-21; B) Uma unidade de repetição dos satélites alinhadas.

O alinhamento 6 (Figura 10) possui satélites descritos em duas espécies, CgoSat06-60 e PliSat05-45, apresentando uma similaridade de aproximadamente 91% entre si, mas uma identidade de apenas 70%, devido a uma porção adicional de 15 pares de base presente em *C. gomesi*. O satélite de *C. gomesi* apresentou cópias mapeadas em todas as espécies, com exceção dos dois representantes dos Characidae, com um número extremamente baixo de cópias mapeadas sugerindo a existência residual do satélite na família. Porém, apenas em *C. gomesi* o satélite de 60 pares de base apresentou uma estrutura em tandem uniforme, enquanto em todas as outras, o *gap* de 15 pares de base sem cópias mapeadas ficou bastante evidente (Figura 11).

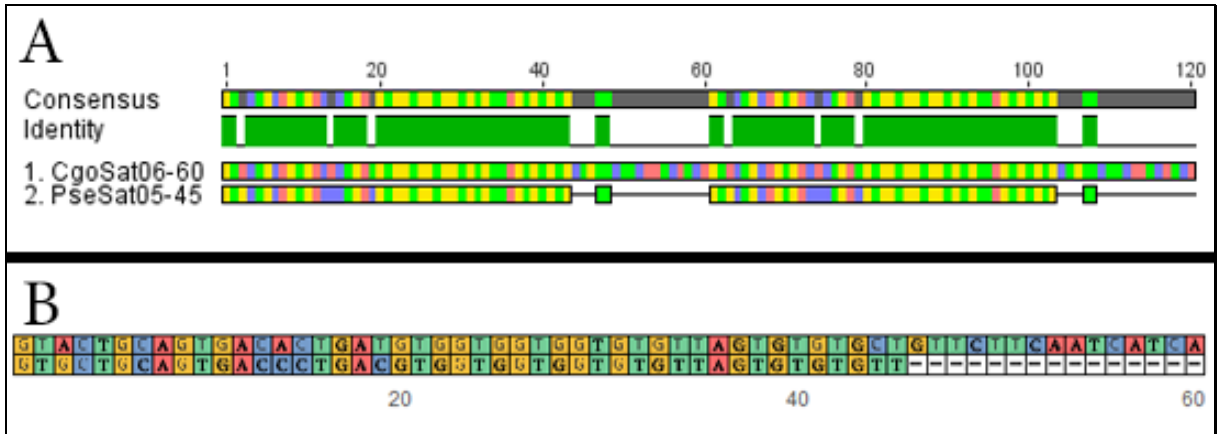


Figura 10: Alinhamento 6: A) Alinhamento de 2 monômeros em tandem dos satélites CgoSat06-60 e PseSat05-45. B) Alinhamento dos monômeros dos satélites.

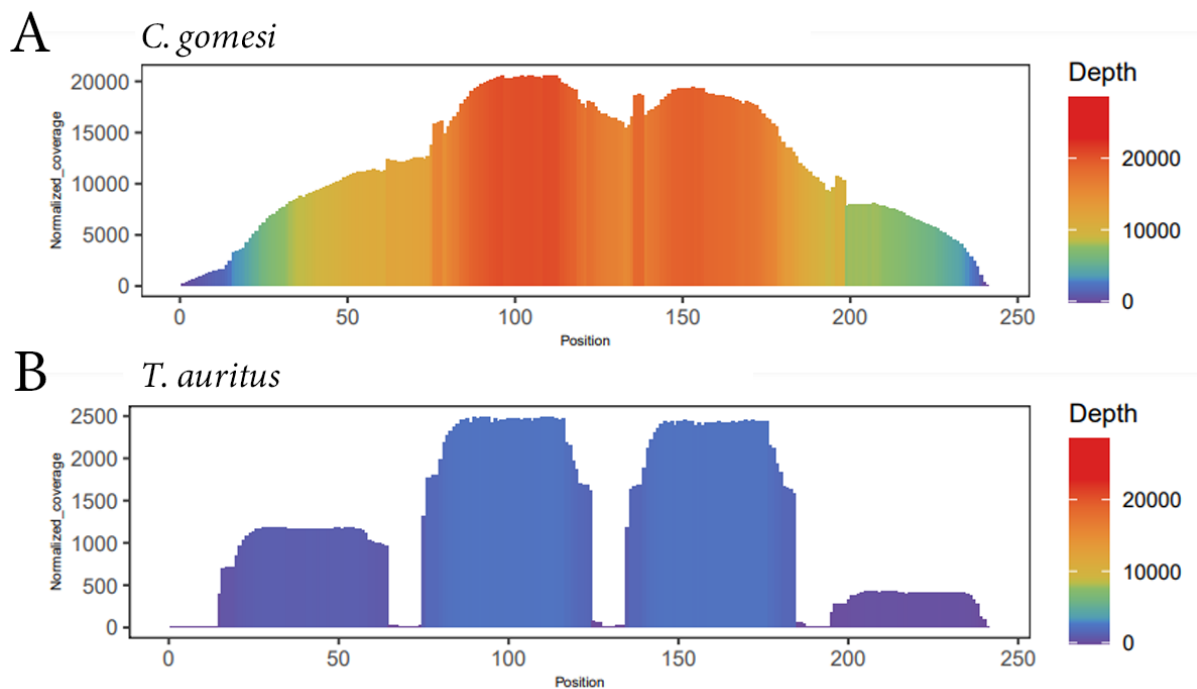


Figura 11: Exemplos dos gráficos de cobertura do CgoSat06-60 produzidos com o RepeatProfiler: A) Gráfico da biblioteca de *C. gomesi*, com uma alta cobertura do satélite, demonstrando a organização em tandem dos arrays; B) Gráfico da biblioteca de *T. auritus*, com uma cobertura considerável, mas irregular, destacando os gaps sem cópias mapeadas.

Finalmente, o alinhamento 7 (Figura 12) revela a ocorrência de 7 satélites com tamanhos que variam de 50 à 54 pares de base, dentre eles CgoSat130-50, que apresenta uma identidade em comparação aos outros satélites do contig que varia de 66 (MasSat08-54) a 78% (LfrSat25-54). Todos os outros satélites do alinhamento apresentam uma identidade

entre si de pelo menos 80%. A análise do perfil do CgoSat130-50 apresentou um número alto de cópias na maioria das espécies, especialmente nos representantes da família Serrasalminidae (com destaque para *P. mesopotamicus*), demonstrando uma organização dos arrays em tandem. Em *C. modestus*, *H. malabaricus*, *P. paranae* e em *P. lineatus* o número de cópias mapeadas foi baixo, mas o gráfico também indica uma estrutura em tandem dos arrays. Em *A. mexicanus* e *T. auritus*, o número de cópias mapeadas foi bastante reduzido, indicando apenas uma presença resquicial do satélite nestas espécies. *P. semifasciata*, por sua vez, não apresentou cobertura no gráfico. Rodrigues (2021) também mapeou o satélite PmeSat21-54 nos cromossomos de *P. mesopotamicus* por meio de FISH, onde foram observados sinais, e portanto organização em *clusters*, nas regiões teloméricas de um par cromossômico da espécie.

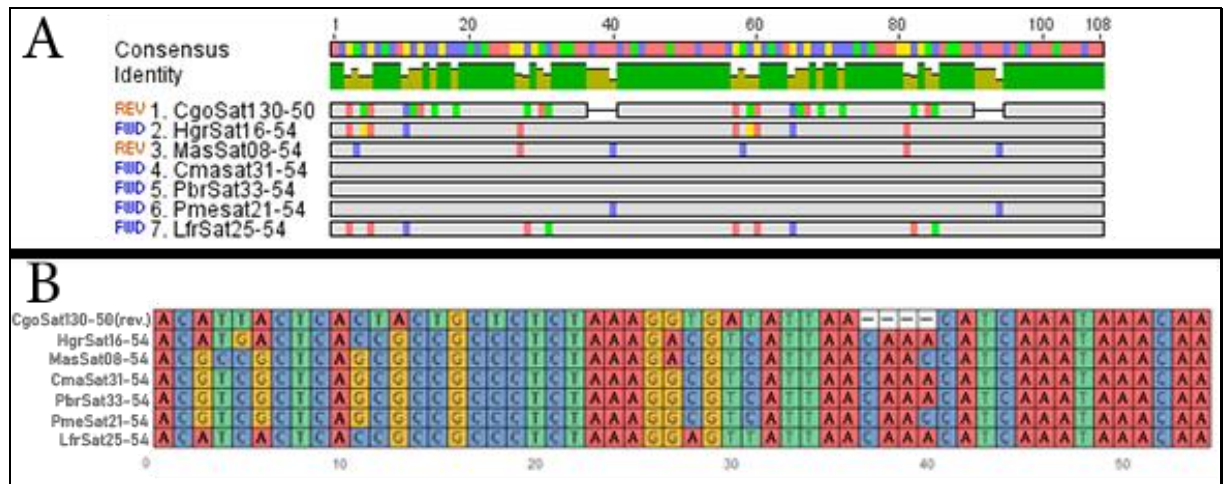


Figura 12: Alinhamento 7: A) Alinhamento de 2 monômeros em tandem dos satélites que apresentam alta similaridade com CgoSat130-50; B) As unidades de repetição de cada um dos satélite alinhadas.

5. DISCUSSÃO

5.1 Caracterização de satelitomas e geração de recursos genômicos

Neste estudo, caracterizamos cinco novos satelitomas de peixes Characoidei (ordem Characiformes), bem como atualizamos e ampliamos o catálogo de *C. gomesi*, totalizando 432 DNAs satélites descritos pela primeira vez, expandindo significativamente o conhecimento acerca da diversidade e abundância destas sequências no genoma de peixes Neotropicais. Até

o momento, este é o primeiro estudo de caracterização e comparação de satelitomas na extensão de toda uma subordem de peixes, que se diversificou há aproximadamente 100 milhões de anos (MELO et al., 2022). O alto número de satélites presentes nos genomas destes peixes está de acordo com o esperado para organismos eucariotos, o que também foi observado em *Drosophila* (MELTERS et al., 2013; DE LIMA; RUIZ-RUANO, 2022), reforçando que o surgimento e a ocorrência de eventos de expansão dos arranjos de satélites são comuns nos genomas destes organismos.

Nossos resultados demonstram que os DNAs satélites representam uma porção genômica bastante variada nas espécies analisadas, inclusive entre espécies pertencentes a famílias mais proximamente relacionadas (e.g. membros de Characidae), apontando para uma rápida evolução destas sequências em peixes Characiformes, com muitos eventos de expansão e contração dos arranjos dos DNAs satélites, até em períodos mais curtos de diversificação.

Além desta alta diversidade no conteúdo total de satélites nos genomas das espécies analisadas, observamos que diferentes satélites predominam nos genomas de cada uma delas. Em muitos casos, os satélites mais abundantes são os mais prováveis candidatos à serem centroméricos, como em alguns gafanhotos (CAMACHO et al., 2022), nos corvos e nas aves do paraíso (PEONA et al., 2022), e em algumas espécies de *Drosophila* (DE LIMA; RUIZ-RUANO, 2022), o que pode ser o caso também para os peixes aqui analisados. Os satélites mais abundantes das espécies analisadas, assim como a maioria dos satélites do estudo, apresentam uma maior proporção de conteúdo A+T, o que segue a tendência observada para outros tipos de organismos como plantas, fungos, e outros vertebrados, mas se diferencia do apresentado pelas aves (PEONA et al., 2022), com uma predominância de satélites mais ricos em G+C. Os dados apresentados contribuem para a realização, no futuro, de experimentos citogenéticos de mapeamento cromossômico e de análises bioinformáticas de investigação da presença de repetições internas nos monômeros dos satélites e da possível formação de estruturas secundárias, indicativos de função, para obtermos mais evidências a respeito dos satélites candidatos ao desempenho de função centromérica. Vale ressaltar que, em aves, nenhum destes outros indicativos foi encontrado para os satélites (PEONA et al., 2022). Em conjunto, esses resultados indicam que a evolução de satélites pode ter uma considerável influência na grande diversidade de espécies apresentada pela ordem.

Após a complementação dos satelitomas de *C. gomesi*, no estudo, e de *P. paranae* (DOS SANTOS, em preparação), observamos que as estimativas da porção dos genomas destas espécies compostas por DNAs satélites era subestimada (SERRANO-FREITAS et al., 2020; SILVA et al., 2017), visto que alguns satélites descritos pela primeira vez nestas

espécies apresentam uma abundância bastante elevada em seus respectivos genomas (como CgoSat60-2063 e ApaSat46-39). Estes resultados revelam que a ferramenta TAREAN (NOVÁK et al., 2017) parece ter maior eficiência na caracterização de satDNAs, embora estudos mais sistematizados devam ser realizados nesse sentido.

A descrição extensiva de satélites realizada aqui também servirá de subsídio na futura anotação genômica desta classe de sequências em todas as espécies do estudo, o que pode ser extrapolado para outras espécies e grupos de peixes, principalmente organismos não-modelo (BOMAN et al., 2019; PEONA et al., 2022). Devido aos DNAs satélites serem extremamente repetitivos, e pode-se dizer homogêneos dentro do genoma das espécies, essas regiões acabam causando problemas durante a montagem de genomas, resultando em uma porção representativa subestimada (em alguns casos formando *contigs* totalmente independentes) ou inexistente, como no genoma anteriormente montado de *A. mexicanus* (MCGAUGH et al., 2014; SILVA et al., 2017). O estudo detalhado da estrutura da organização dos arranjos dos DNAs satélites só é possível com a análise de *reads* longos ou genomas montados, o que destaca a importância da existência de sequências referências de satélites para auxiliar na anotação desta porção repetitiva em diferentes espécies.

5.2 Evolução e compartilhamento de satélites conservados

Os sete alinhamentos demonstrados aqui representam satélites descritos que apresentam uma organização em tandem, como evidenciado pelos gráficos gerados na análise do TAREAN, e também pelos gráficos do RepeatProfiler. Devido à extensão destes satélites em diferentes espécies, sugerimos que todos se originaram, pelo menos, no último ancestral em comum de toda a subordem Characoidei, com idade estimada em aproximadamente 140 a 78 milhões de anos, antes da separação de Crenuchidae com as outras famílias do grupo (MELO et al., 2022). Após se originar, estes satélites se mantiveram, em sua maioria, extremamente conservados durante a história evolutiva das linhagens estudadas. Assim como no caso do CharSat01-52, os satDNAs aqui caracterizados estão entre os mais antigos já descritos, em conjunto com outros satélites de moluscos (BIV160, por exemplo) e de *Drosophila* (tapiR), com idades estimadas de aproximadamente 540 e 200 milhões de anos, respectivamente (PLOHL et al., 2010; PETRACCIOLI et al., 2015; HALBACH et al., 2020; DOS SANTOS et al., 2021).

Seguindo o exemplo de Dos Santos et al. (2021), sugerimos os nomes CharSat02-21 para CgoSat71-21, CharSat03-29 para CgoSat26-29, CharSat04-42 para CgoSat10-42,

CharSat05-21 para CgoSat66-21, CharSat06-21 para CgoSat32-21, CharSat07-60 para CgoSat06-60, e CharSat08-50 para CgoSat130-50, indicando que são DNAs satélites compartilhados por toda a subordem Characoidei.

Atualmente, considera-se que os modelos da hipótese *library* em conjunto com a evolução em concerto podem explicar a evolução de uma grande maioria dos DNAs satélites descritos até então. Isso se aplica aos resultados aqui encontrados, onde as espécies apresentaram satelitomas com uma composição bastante diversa, onde a maioria dos satélites passaram por eventos de diferenciação ao ponto de se descaracterizarem em relação às famílias de DNAs satélites presentes em um ancestral em comum, mas ao mesmo tempo compartilhando uma pequena porção de famílias de satélites altamente conservadas que tiveram origem antes da diversificação das espécies estudadas, e que hoje apresentam apenas alterações quantitativas. De um total de 981 satélites estudados, a maioria são específicos, e apenas sete satélites se encontram altamente conservados no genoma de espécies distribuídas pelas quatro superfamílias de Characoidei estudadas (Crenuchoidea, Erythrinoidea, Curimatoidea e Characoidea). O satélite CharSat01-52 (DOS SANTOS et al., 2021) também se manteve altamente conservado em Crenuchoidea, Curimatoidea e Characoidea, além de Alestoidea (superfamília dos peixes da ordem que estão distribuídos pelo continente africano, com nenhuma amostra genômica analisada aqui), mas apresentou apenas resquícios de sua existência no genoma de *H. malabaricus* (Erythrinoidea). Esses oito casos de conservação contrariam as predições de evolução molecular de satDNAs.

Nossos resultados indicam que o processo de homogeneização vem acontecendo de maneira similar nos arranjos dos satélites presentes em espécies que divergiram há muito tempo, mais especificamente, em toda uma subordem. O alinhamento 4 mostra a composição e organização do satélite CgoSat66-21 em *C. macropomum* como uma HOR da variante descrita nas outras espécies, um exemplo claro de diferenciação inicial das variantes nas diferentes espécies relacionadas, que está de acordo com a hipótese *library* e com o modelo de evolução de DNAs satélites proposto por Ruiz-Ruano et al. (2016), e também demonstra a homogeneização acontecendo nos *arrays* do satélite em *C. macropomum*, sofrendo os mesmos efeitos da evolução em concerto.

No alinhamento 6, do CgoSat06-60, a diferença no perfil do gráfico de cobertura do satélite apresentado pelas espécies indica a presença do satélite com 60 pares de base em *C. gomesi*, e a presença do satélite com 45 pares de base (praticamente idêntica à maior porção do satélite de *C. gomesi*) em todas as outras bibliotecas analisadas. Ou a variante deste satélite que dominou o genoma de *C. gomesi* já possuía 60 pares de base, passando pela deleção de

uma porção de 15 pares de base nas outras espécies, ou o satélite compartilhado por todas as espécies possuía 45 pares de base inicialmente, se caracterizando na nova organização de 60 pares de base no Crenuchidae posteriormente, após a acumulação de eventos de diferenciação desta sequência. Novamente, este resultado é um exemplo de compartilhamento de satélites por espécies relacionadas, com uma abundância diferencial entre elas, exemplificando o modelo da hipótese *library*. Também demonstra a atuação da evolução em concerto nos arranjos das espécies, indicando uma homogeneização diferencial intraespecífica do satélite em *C. gomesi*, ou uma homogeneização diferencial em todas as outras espécies.

A alta conservação da composição dos satélites por toda uma ordem ou subordem desafia as previsões do modelo da evolução em concerto, na qual um satDNA evolui rapidamente e de maneira independente, invariavelmente levando à diferenciação interespecífica ao longo do tempo evolutivo, o que é observado na maioria dos satélites descritos, que já se diferenciaram em algum grau dentro do genoma de cada espécie. Por esta razão, casos de elevada conservação não são esperados e levantam dúvidas sobre as possíveis causas destas ocorrências, o que inclui a possibilidade da existência de função por parte destas sequências, por exemplo (BISCOTTI; OLMO; HESLOP-HARRISON, 2015; DOS SANTOS et al., 2021).

Outros fatores que poderiam influenciar na conservação da composição do satélite por longos períodos seria o fato de apresentarem combinações de nucleotídeos que favorecessem a atuação dos mecanismos de homogeneização dos mesmos (PLOHL et al. 2008; 2012; DOS SANTOS et al., 2021). Também, uma lenta taxa de evolução (baixa taxa de mutação e de homogeneização) poderia explicar essa conservação, como no caso dos esturjões (DE LA HERRÁN et al., 2001), ou na possibilidade levantada para a manutenção do CharSat01-52 em Characoidei por Dos Santos et al. (2021). Nossos dados não corroboram esta hipótese para os satélites dos peixes aqui estudados, visto que de um total de 981 satélites analisados, apenas oito casos (incluindo o CharSat01-52) de alta conservação da composição nucleotídica ao longo da subordem foram identificados, o que não indica uma lenta evolução da porção repetitiva composta por DNAs satélites dos genomas dessas espécies.

Fry e Salser (1977) também sugeriram que uma das etapas do percurso evolutivo de um DNA satélite é a aquisição de função, o que não descarta a possibilidade da manutenção de um satDNA durante um longo período evolutivo independentemente dos mecanismos da seleção natural (CAMACHO et al., 2022). Por exemplo, ncRNA derivados de satélites podem estar envolvidos em diversas funções celulares. Se em futuras análises com bibliotecas de sequenciamento de RNA destas espécies forem encontrados sinais de transcrição dos satélites

conservados descritos aqui, assim como foi encontrado para CharSat01-52 (DOS SANTOS et al., 2021), essa seria uma possível explicação para a sua manutenção ao longo de milhões de anos de evolução (PEZER et al., 2012; PETRACCIOLI et al., 2015; FERREIRA et al., 2019; HALBACH et al., 2020; LOUZADA et al., 2020; DOS SANTOS ET AL., 2021).

Se estes satélites estão sendo mantidos conservados devido à seleção por apresentarem alguma função genômica e consequentemente efeitos no fitness dos organismos é uma questão a ser investigada no futuro, por meio de análises de localização dos arranjos próximos à genes, na busca por *motifs* funcionais ou de formação de estruturas secundárias nas sequências dos satélites (DOS SANTOS et al., 2021; CAMACHO et al., 2022).

Também, análises em genomas montados ou em *reads* longos podem nos ajudar a verificar a organização dos arranjos dos satélites nas espécies, permitindo analisar a ação da evolução em concerto nestas sequências, e se estes demonstram diferentes organizações, como no caso dos genomas dos corvos e das aves do paraíso (PEONA et al., 2022), além de verificar a possível associação com elementos transponíveis (KHOST; EICKBUSH; LARRACUENTE, 2017; CECHOVA et al., 2019; HEITKAM et al., 2020; VONDRAK et al., 2020).

5.3 Considerações finais

Ainda que diversos estudos envolvendo os DNAs satélites em peixes Neotropicais venham sendo desenvolvidos (SILVA et al., 2017; UTSUNOMIA et al., 2019; CREPALDI; PARISE-MALTEMPI, 2020; SERRANO-FREITAS et al., 2020; STORNIOLI et al., 2021; KRETCHSMER et al., 2022), nosso entendimento a respeito da evolução de satélites nesse grupo ainda é limitado, especialmente devido à alta diversidade de espécies que apresenta.

Para a maioria das espécies ainda não existem genomas montados, o que é ainda mais problemático se tratando da porção repetitiva dos genomas, que inclui os arranjos dos DNAs satélites. Nestes casos, é importante realizar estudos que combinem a aplicação de diferentes tecnologias na busca do entendimento da localização cromossômica e da organização desses arranjos no genoma das espécies, e para isso, combinados às análises bioinformáticas, os estudos citogenéticos se tornam essenciais (DEAKIN et al., 2019; DOS SANTOS et al., 2021; DE LIMA; RUIZ-RUANO, 2022; PEONA et al., 2022).

Neste estudo caracterizamos cinco novos satelitomas de peixes Characoidei (ordem Characiformes) e atualizamos outro catálogo previamente descrito, expandindo assim o conhecimento acerca da composição e de alguns traços destas sequências no genoma das

espécies aqui estudadas. Além disto, identificamos sete famílias de DNA satélite que são compartilhadas por toda a subordem Characoidei. Os dados aqui obtidos abrem portas para análises citogenéticas de mapeamento cromossômico destas sequências, análises de transcrição, e análises bioinformáticas de localização e colocalização dos *arrays* dos satélites em *reads* longos ou genomas montados, em busca de evidências sobre o motivo da manutenção destas sequências altamente conservadas no genoma de diferentes espécies, desafiando as previsões da evolução em concerto em um contexto de ausência de função.

6. REFERÊNCIAS

- ALEXANDROV, Ivan et al. Alpha-satellite DNA of primates: old and new families. **Chromosoma**, v. 110, n. 4, p. 253-266, 2001.
- AMARAL, Paulo P.; MATTICK, John S. Noncoding RNA in development. **Mammalian genome**, v. 19, n. 7, p. 454-492, 2008.
- ARCILA, Dahiana et al. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. **Nature Ecology & Evolution**, v. 1, n. 2, p. 1-10, 2017.
- BETANCUR-R, Ricardo et al. The tree of life and a new classification of bony fishes. **PLoS currents**, v. 5, 2013.
- BETANCUR-R, Ricardo et al. Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes. **Evolution**, v. 73, n. 2, p. 329-345, 2019.
- BISCOTTI, Maria Assunta; OLMO, Ettore; HESLOP-HARRISON, J. S. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415-420, 2015.
- BOMAN, Jesper et al. The genome of blue-capped cordon-bleu uncovers hidden diversity of LTR retrotransposons in zebra finch. **Genes**, v. 10, n. 4, p. 301, 2019.
- BOTCHAN, Michael R. Bovine satellite I DNA consists of repetitive units 1,400 base pairs in length. **Nature**, v. 251, n. 5473, p. 288-292, 1974.
- BURNS, Michael D.; SIDLAUSKAS, Brian L. Ancient and contingent body shape diversification in a hyperdiverse continental fish radiation. **Evolution**, v. 73, n. 3, p. 569-587, 2019.
- CAMACHO, Juan Pedro M. et al. Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. **BMC biology**, v. 20, n. 1, p. 1-24, 2022.
- CECHOVA, Monika et al. High satellite repeat turnover in great apes studied with short-and long-read technologies. **Molecular biology and evolution**, v. 36, n. 11, p. 2415-2431, 2019.

- 767 CHAKRABARTY, Prosanta et al. Phylogenomic systematics of ostariophysan fishes:
768 ultraconserved elements support the surprising non-monophyly of Characiformes. **Systematic**
769 **Biology**, v. 66, n. 6, p. 881-895, 2017.
- 770 CHARLESWORTH, Brian; SNIEGOWSKI, Paul; STEPHAN, Wolfgang. The evolutionary
771 dynamics of repetitive DNA in eukaryotes. **Nature**, v. 371, n. 6494, p. 215-220, 1994.
- 772 CHEN, Wei-Jen; LAVOUÉ, Sébastien; MAYDEN, Richard L. Evolutionary origin and early
773 biogeography of otophysan fishes (Ostariophysi: Teleostei). **Evolution**, v. 67, n. 8, p. 2218-
774 2239, 2013.
- 775 CREPALDI, Carolina; PARISE-MALTEMPI, Patricia P. Heteromorphic sex chromosomes
776 and their DNA content in fish: an insight through satellite DNA accumulation in
777 *Megaleporinus elongatus*. **Cytogenetic and Genome Research**, v. 160, n. 1, p. 38-46, 2020.
- 778 DAI, Wei et al. Phylogenomic perspective on the relationships and evolutionary history of the
779 major otocephalan lineages. **Scientific reports**, v. 8, n. 1, p. 1-12, 2018.
- 780 DE LA HERRÁN, Roberto et al. Slow rates of evolution and sequence homogenization in an
781 ancient satellite DNA family of sturgeons. **Molecular Biology and Evolution**, v. 18, n. 3, p.
782 432-436, 2001.
- 783 DEAKIN, Janine E. et al. Chromosomics: Bridging the gap between genomes and
784 chromosomes. **Genes**, v. 10, n. 8, p. 627, 2019.
- 785 DE LIMA, Leonardo G.; RUIZ-RUANO, Francisco J. In-depth Satellitome Analyses of 37
786 *Drosophila* species illuminate repetitive DNA evolution in the *Drosophila* genus. **Genome**
787 **Biology and Evolution**, v. 14, n. 5, p. evac064, 2022.
- 788 DIMMICK, Walter W.; LARSON, Allan. A molecular and morphological perspective on the
789 phylogenetic relationships of the otophysan fishes. **Molecular phylogenetics and evolution**,
790 v. 6, n. 1, p. 120-133, 1996.
- 791 DOS SANTOS, Rodrigo Z. et al. A long-term conserved satellite DNA that remains
792 unexpanded in several genomes of Characiformes fish is actively transcribed. **Genome**
793 **biology and evolution**, v. 13, n. 2, p. evab002, 2021.
- 794 DOS SANTOS, Rodrigo Z. **Estrutura, organização e atividade de DNAs satélites no**
795 **genoma do lambari *Astyanax paranae* (Teleostei, Characiformes)**. UNESP, em
796 elaboração.
- 797 DOVER, Gabriel A. Molecular drive in multigene families: how biological novelties arise,
798 spread and are assimilated. **Trends in genetics**, v. 2, p. 159-165, 1986.
- 799 FERREIRA, Daniela et al. DNA methylation patterns of a satellite non-coding sequence-FA-
800 SAT in cancer cells: Its expression cannot be explained solely by DNA methylation.
801 **Frontiers in genetics**, v. 10, p. 101, 2019.
- 802 FINK, Sara V.; FINK, William L. Interrelationships of the ostariophysan fishes
803 (Teleostei). **Zoological Journal of the Linnean Society**, v. 72, n. 4, p. 297-353, 1981.

- 804 FRICKE, Ron. **ESCHMEYER'S CATALOG OF FISHES: GENERA, SPECIES,**
 805 **REFERENCES.** 2021. Disponível em:
 806 <https://www.calacademy.org/scientists/projects/eschmeyers-catalog-of-fishes>. Acesso em: 20
 807 fev. 2021.
- 808 FRY, Kirk; SALSER, Winston. Nucleotide sequences of HS- α satellite DNA from kangaroo
 809 rat *Dipodomys ordii* and characterization of similar sequences in other rodents. **Cell**, v. 12, n.
 810 4, p. 1069-1084, 1977.
- 811 GARRIDO-RAMOS, Manuel A. Satellite DNA in plants: more than just
 812 rubbish. **Cytogenetic and Genome Research**, v. 146, n. 2, p. 153-170, 2015.
- 813 GARRIDO-RAMOS, Manuel A. Satellite DNA: an evolving topic. **Genes**, v. 8, n. 9, p. 230,
 814 2017.
- 815 GOES, Caio A. G. et al. Cytogenetic markers as a tool for characterization of hybrids of
 816 *Astyanax Baird & Girard*, 1854 and *Hyphessobrycon Eigenmann*, 1907. **Comparative**
 817 **Cytogenetics**, v. 14, n. 2, p. 231, 2020.
- 818 HALBACH, Rebecca et al. A satellite repeat-derived piRNA controls embryonic
 819 development of *Aedes*. **Nature**, v. 580, n. 7802, p. 274-277, 2020.
- 820 HASHIMOTO, Diogo T. et al. Molecular diagnostic methods for identifying Serrasalmid fish
 821 (*Pacu*, *Pirapitinga*, and *Tambaqui*) and their hybrids in the Brazilian aquaculture
 822 industry. **Aquaculture**, v. 321, n. 1-2, p. 49-53, 2011.
- 823 HEITKAM, Tony et al. Satellite DNA landscapes after allotetraploidization of quinoa
 824 (*Chenopodium quinoa*) reveal unique A and B subgenomes. **The Plant Journal**, v. 103, n. 1,
 825 p. 32-52, 2020.
- 826 HÖRZ, Wolfram; ZACHAU, Hans G. Characterization of distinct segments in mouse satellite
 827 DNA by restriction nucleases. **European Journal of Biochemistry**, v. 73, n. 2, p. 383-392,
 828 1977.
- 829 HUTTON, James R.; WETMUR, James G. Effect of chemical modification on the rate of
 830 renaturation of deoxyribonucleic acid. Deaminated and glyoxalated deoxyribonucleic
 831 acid. **Biochemistry**, v. 12, n. 3, p. 558-563, 1973.
- 832 HUNTER, John D. Matplotlib: A 2D graphics environment. **Computing in science &**
 833 **engineering**, v. 9, n. 03, p. 90-95, 2007. KEARSE, M. et al. Geneious Basic: An integrated
 834 and extendable desktop software platform for the organization and analysis of sequence data.
 835 **Bioinformatics**, v. 28, n. 12, p. 1647-1649, 2012.
- 836 KHOST, Daniel E.; EICKBUSH, Danna G.; LARRACUENTE, Amanda M. Single-molecule
 837 sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila*
 838 *melanogaster*. **Genome research**, v. 27, n. 5, p. 709-721, 2017.
- 839 KOPECKA, Helena et al. Restriction enzyme analysis of satellite DNA components from the
 840 bovine genome. **European journal of biochemistry**, v. 84, n. 1, p. 189-195, 1978.

- 841 KRETSCHMER, Rafael et al. Satellitome analysis illuminates the evolution of ZW sex
842 chromosomes of Triportheidae fishes (Teleostei: Characiformes). **Chromosoma**, p. 1-17,
843 2022.
- 844 KUHN, Gustavo C. S. Satellite DNA transcripts have diverse biological roles in
845 *Drosophila*. **Heredity**, v. 115, n. 1, p. 1, 2015.
- 846 LANGMEAD, Ben; SALZBERG, Steven L. Fast gapped-read alignment with Bowtie
847 2. **Nature methods**, v. 9, n. 4, p. 357-359, 2012.
- 848 LI, Heng et al. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n.
849 16, p. 2078-2079, 2009.
- 850 LORITE, Pedro et al. Concerted evolution, a slow process for ant satellite DNA: study of the
851 satellite DNA in the *Aphaenogaster* genus (Hymenoptera, Formicidae). **Organisms Diversity**
852 **& Evolution**, v. 17, n. 3, p. 595-606, 2017.
- 853 LOUZADA, Sandra et al. Decoding the role of satellite DNA in genome architecture and
854 plasticity—An evolutionary and clinical affair. **Genes**, v. 11, n. 1, p. 72, 2020.
- 855 LOWER, Sarah S. et al. Satellite DNA evolution: old ideas, new approaches. **Current**
856 **opinion in genetics & development**, v. 49, p. 70-78, 2018.
- 857 MCGAUGH, Suzanne E. et al. The cavefish genome reveals candidate genes for eye loss.
858 **Nature communications**, v. 5, n. 1, p. 1-10, 2014.
- 859 MELO, Bruno F. et al. Accelerated diversification explains the exceptional species richness of
860 tropical characoid fishes. **Systematic Biology**, v. 71, n. 1, p. 78-92, 2022.
- 861 MELTERS, Daniël P. et al. Comparative analysis of tandem repeats from hundreds of species
862 reveals unique insights into centromere evolution. **Genome biology**, v. 14, n. 1, p. 1-20, 2013.
- 863 MIGA, Karen H. et al. Telomere-to-telomere assembly of a complete human X
864 chromosome. **Nature**, v. 585, n. 7823, p. 79-84, 2020.
- 865 MOREIRA, Cristiano R. et al. **Relações filogenéticas na ordem Characiformes (Teleostei:**
866 **Ostariophysi)**. Universidade de São Paulo, v. 468, 2007.
- 867 NAKATANI, Masanori et al. Evolutionary history of Otophysi (Teleostei), a major clade of
868 the modern freshwater fishes: Pangaeian origin and Mesozoic radiation. **BMC evolutionary**
869 **biology**, v. 11, n. 1, p. 1-25, 2011.
- 870 NEAR, Thomas J. et al. Resolution of ray-finned fish phylogeny and timing of
871 diversification. **Proceedings of the National Academy of Sciences**, v. 109, n. 34, p. 13698-
872 13703, 2012.
- 873 NEGM, Sherif et al. RepeatProfiler: a pipeline for visualization and comparative analysis of
874 repetitive DNA profiles. **Molecular ecology resources**, v. 21, n. 3, p. 969-981, 2021.
- 875 NOVÁK, Petr et al. RepeatExplorer: A Galaxy-based web server for genome-wide
876 characterization of eukaryotic repetitive elements from next-generation sequence reads.
877 **Bioinformatics**, v. 29, n. 6, p. 792–793, 2013.

- 878 NOVÁK, Petr et al. TAREAN: A computational tool for identification and characterization of
879 satellite DNA from unassembled short reads. **Nucleic Acids Research**, v. 45, n. 12, 2017.
- 880 OLIVEIRA, Claudio et al. Phylogenetic relationships within the speciose family Characidae
881 (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup
882 sampling. **BMC evolutionary biology**, v. 11, n. 1, p. 1-25, 2011.
- 883 PEONA, Valentina et al. Satellite DNA evolution in Corvoidea inferred from short and long
884 reads. **Molecular Ecology**, 2022.
- 885 PETRACCIOLI, Agnese et al. A novel satellite DNA isolated in *Pecten jacobaeus* shows high
886 sequence similarity among molluscs. **Molecular Genetics and Genomics**, v. 290, n. 5, p.
887 1717-1725, 2015.
- 888 PEZER, Željka et al. Satellite DNA-mediated effects on genome regulation. **Repetitive DNA**,
889 v. 7, p. 153-169, 2012.
- 890 PLOHL, Miroslav et al. Long-term conservation vs high sequence divergence: the case of an
891 extraordinarily old satellite DNA in bivalve mollusks. **Heredity**, v. 104, n. 6, p. 543-551,
892 2010.
- 893 PLOHL, Miroslav; MEŠTROVIĆ, Nevenka; MRAVINAC, Brankica. Satellite DNA
894 evolution. **Repetitive DNA**, v. 7, p. 126-152, 2012.
- 895 PORTO-FORESTI, Fábio et al. Cytogenetic markers as diagnoses in the identification of the
896 hybrid between Piaçu (*Leporinus macrocephalus*) and Piapara (*Leporinus*
897 *elongatus*). **Genetics and Molecular Biology**, v. 31, n. 1, p. 195-202, 2008.
- 898 RODRIGUES, Pedro H. M. **Caracterização e análises citogenômicas de satelitomas de**
899 **Serrasalmideos (Teleostei, Characiformes)**. 2021. 55 f. Dissertação (Mestrado Acadêmico
900 em Biociências) – Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras,
901 Assis, 2021.
- 902 RUIZ-RUANO, Francisco J. et al. High-throughput analysis of the satellitome illuminates
903 satellite DNA evolution. **Scientific reports**, v. 6, n. 1, p. 1-14, 2016.
- 904 RUIZ-RUANO, Francisco J. et al. High-throughput analysis of satellite DNA in the
905 grasshopper *Pyrgomorpha conica* reveals abundance of homologous and heterologous higher-
906 order repeats. **Chromosoma**, v. 127, n. 3, p. 323-340, 2018.
- 907 SCHMIEDER, Robert; EDWARDS, Robert. Fast identification and removal of sequence
908 contamination from genomic and metagenomic datasets. **PloS one**, v. 6, n. 3, p. e17288, 2011.
- 909 SERRANO-FREITAS, Érica A. et al. Satellite DNA content of B chromosomes in the
910 characid fish *Characidium gomesi* supports their origin from sex chromosomes. **Molecular**
911 **Genetics and Genomics**, v. 295, n. 1, p. 195-207, 2020.
- 912 SILVA, Duilio M. Z. A. et al. High-throughput analysis unveils a highly shared satellite DNA
913 library among three species of fish genus *Astyanax*. **Scientific Reports**, v. 7, n. 1, p. 1-12,
914 2017.

- 915 SMALEC, Brendan M. et al. A centromere satellite concomitant with extensive karyotypic
916 diversity across the *Peromyscus* genus defies predictions of molecular drive. **Chromosome**
917 **Research**, v. 27, n. 3, p. 237-252, 2019.
- 918 SMIT, Arian F. A. Repeat-Masker Open-3.0. <http://www.repeatmasker.org>, 2004.
- 919 SMITH, George P. Evolution of Repeated DNA Sequences by Unequal Crossover: DNA
920 whose sequence is not maintained by selection will develop periodicities as a result of random
921 crossover. **Science**, v. 191, n. 4227, p. 528-535, 1976.
- 922 STORNIOLI, José Henrique F. et al. The B chromosomes of *Prochilodus lineatus* (Teleostei,
923 Characiformes) are highly enriched in satellite DNAs. **Cells**, v. 10, n. 6, p. 1527, 2021.
- 924 TAYLOR, Richard. Interpretation of the correlation coefficient: a basic review. **Journal of**
925 **diagnostic medical sonography**, v. 6, n. 1, p. 35-39, 1990.
- 926 TEAM, R. Core et al. R: A language and environment for statistical computing. 2013.
- 927 TUTAR, Yusuf. Pseudogenes. **Comparative and functional genomics**, v. 2012, 2012.
- 928 URALSKY, Lev I. et al. Classification and monomer-by-monomer annotation dataset of
929 suprachromosomal family 1 alpha satellite higher-order repeats in hg38 human genome
930 assembly. **Data in brief**, v. 24, p. 103708, 2019.
- 931 UTSUNOMIA, Ricardo et al. Genetic differentiation among distinct karyomorphs of the wolf
932 fish *Hoplias malabaricus* species complex (Characiformes, Erythrinidae) and report of
933 unusual hybridization with natural triploidy. **Journal of Fish Biology**, v. 85, n. 5, p. 1682-
934 1692, 2014.
- 935 UTSUNOMIA, Ricardo et al. Satellitome landscape analysis of *Megaleporinus*
936 *macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences
937 on the heteromorphic sex chromosome. **Scientific reports**, v. 9, n. 1, p. 1-10, 2019.
- 938 VAN ROSSUM, Guido; DRAKE, Fred L. Python 3 reference manual. **Scotts Valley, CA:**
939 **createspace**, 2009.
- 940 VICARI, Marcelo R. et al. Satellite DNA and chromosomes in Neotropical fishes: methods,
941 applications and perspectives. **Journal of Fish Biology**, v. 76, n. 5, p. 1094-1116, 2010.
- 942 VONDRAK, Tihana et al. Characterization of repeat arrays in ultra-long nanopore reads
943 reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. **The**
944 **Plant Journal**, v. 101, n. 2, p. 484-500, 2020.
- 945 WASKOM, Michael L. Seaborn: statistical data visualization. **Journal of Open Source**
946 **Software**, v. 6, n. 60, p. 3021, 2021.
- 947 WILLARD, Huntington F. Evolution of alpha satellite. **Current opinion in genetics &**
948 **development**, v. 1, n. 4, p. 509-514, 1991.
- 949
- 950
- 951 **7. MATERIAL SUPPLEMENTAR**

952

953 **Tabela Suplementar 1:** As características principais dos satélites de *H. malabaricus*.

Satélite	RUL	Divergência	Abundância	A+T(%)
HmaSat01-513	513	3,91	0,021349729	60,4
HmaSat02-58	58	3,5	0,017459219	41,4
HmaSat03-326	326	4,91	0,014004296	62,9
HmaSat04-280	280	4,34	0,013427051	61,8
HmaSat05-613	613	2,55	0,008278522	50,2
HmaSat06-46	46	10,35	0,003311804	47,8
HmaSat07-192	192	3,35	0,002320505	64,1
HmaSat08-1750	1750	15,59	0,002118086	61,1
HmaSat09-325	325	5,44	0,00198271	57,8
HmaSat10-187	187	11,55	0,001981092	69
HmaSat11-673	673	2,06	0,001457462	49,3
HmaSat12-30	30	4,69	0,001423077	43,3
HmaSat13-967	967	1,56	0,00127414	64,3
HmaSat14-2774	2774	7,85	0,000981434	65,2
HmaSat15-58	58	3,65	0,000777821	51,7
HmaSat16-168	168	6,34	0,000682331	71,4
HmaSat17-152	152	11,74	0,000664605	63,8
HmaSat18-31	31	13,59	0,000647464	48,4
HmaSat19-43	43	4,59	0,000628207	53,5
HmaSat20-49	49	5,65	0,00058965	46,9
HmaSat21-322	322	10,17	0,000539633	75,8
HmaSat22-1040	1040	2,53	0,000506898	52,6
HmaSat23-42	42	5,64	0,000433703	52,4
HmaSat24-142	142	9,7	0,000431226	63,4
HmaSat25-972	972	3,84	0,000395658	45,9
HmaSat26-910	910	1,91	0,000337998	49,9
HmaSat27-237	237	12,39	0,000323374	41,8
HmaSat28-28	28	8,72	0,000318537	67,9
HmaSat29-51	51	6,72	0,000286963	66,7
HmaSat30-403	403	3,06	0,000284561	56,3
HmaSat31-284	284	8,38	0,000282706	63,4
HmaSat32-695	695	2,46	0,000276422	47,9
HmaSat33-729	729	1,09	0,000276213	50,6
HmaSat34-632	632	3,44	0,000262265	55,4
HmaSat35-1378	1378	0,8	0,00024979	44,1
HmaSat36-1027	1027	1,42	0,000226423	50
HmaSat37-683	683	1,41	0,000216687	49,5
HmaSat38-27	27	9,4	0,000214557	59,3
HmaSat39-767	767	0,65	0,00021221	57,6

HmaSat40-129	129	10,17	0,000195037	71,3
HmaSat41-176	176	12,83	0,000193947	69,3
HmaSat42-1756	1756	2,45	0,000186158	57,6
HmaSat43-39	39	9,55	0,000163019	38,5
HmaSat44-941	941	5,83	0,00015898	59,4
HmaSat45-199	199	5,85	0,000157088	59,8
HmaSat46-760	760	1,66	0,000147378	53,6
HmaSat47-1420	1420	2,47	0,000145402	61,4
HmaSat48-49	49	5,19	0,000140011	46,9
HmaSat49-174	174	8,02	0,000134992	61,5
HmaSat50-21	21	6,8	0,000115041	71,4
HmaSat51-219	219	2,67	0,000112907	53
HmaSat52-211	211	4,78	0,00011133	58,3
HmaSat53-56	56	6,52	0,00010917	53,6
HmaSat54-93	93	8,9	0,000104253	62,4
HmaSat55-6	6	24,56	5,71822E-05	50
HmaSat56-1760	1760	2,05	4,66416E-05	62,3

954

955 **Tabela Suplementar 2:** As características principais dos satélites de *L. friderici*.

Satélite	RUL	Divergência	Abundância	A+T(%)
LfrSat01-48	48	9,36	0,014493218	62,5
LfrSat02-42	42	13,25	0,005557683	57,1
LfrSat03-291	291	8,01	0,004706277	43,6
LfrSat04-694	694	7,11	0,004549426	47,7
LfrSat05-537	537	9,16	0,003905347	61,1
LfrSat06-605	605	6,02	0,003733762	55
LfrSat07-471	471	9,03	0,002281624	56,7
LfrSat08-62	62	4,87	0,002084119	71
LfrSat09-37	37	6,64	0,002069733	56,8
LfrSat10-677	677	9,6	0,001960178	58,1
LfrSat11-178	178	9,95	0,00192396	66,9
LfrSat12-166	166	9,93	0,001642228	53,6
LfrSat13-53	53	8,37	0,001629782	58,5
LfrSat14-1062	1062	5,14	0,001601327	59,9
LfrSat15-965	965	1,7	0,001438129	63,8
LfrSat16-246	246	14,16	0,001375683	61,8
LfrSat17-39	39	8,26	0,00121697	61,15
LfrSat18-67	67	10,62	0,00112405	62,7
LfrSat19-246	246	14,7	0,00108798	62,2
LfrSat20-25	25	8,39	0,001083386	60
LfrSat21-1011	1011	14,08	0,001018604	56,9
LfrSat22-35	35	9,25	0,001002604	51,4

LfrSat23-32	32	11,68	0,000922396	46,9
LfrSat24-22	22	8,2	0,000872089	59,1
LfrSat25-54	54	10,08	0,000855238	61,1
LfrSat26-258	258	6,37	0,00078797	56,2
LfrSat27-29	29	13,71	0,000780198	48,3
LfrSat28-120	120	9,54	0,000752446	52,5
LfrSat29-65	65	7,6	0,000720594	64,6
LfrSat30-245	245	20,79	0,000715297	59,6
LfrSat31-708	708	17,95	0,000677881	55,6
LfrSat32-509	509	4,59	0,000643149	60,1
LfrSat33-39	39	10,92	0,00063805	53,8
LfrSat34-6	6	23,81	0,000623376	50
LfrSat35-24	24	17,09	0,000611436	50
LfrSat36-90	90	5,83	0,000605168	70
LfrSat37-21	21	14,59	0,000590822	71,4
LfrSat38-532	532	17,33	0,000568238	54,5
LfrSat39-1031	1031	3,35	0,000558238	55,9
LfrSat40-34	34	11,47	0,000530891	64,7
LfrSat41-284	284	13,34	0,000519545	56,3
LfrSat42-48	48	7,83	0,000509297	45,8
LfrSat43-30	30	12,69	0,000462178	63,3
LfrSat44-52	52	15,29	0,000449901	65,4
LfrSat45-34	34	7,28	0,000436515	64,7
LfrSat46-41	41	7,09	0,000425099	70,7
LfrSat47-44	44	6,66	0,000415723	63,6
LfrSat48-37	37	7,64	0,000409574	56,8
LfrSat49-707	707	4,21	0,000408644	53,5
LfrSat50-148	148	16,91	0,000399822	60,1
LfrSat51-531	531	10,76	0,000397465	59,1
LfrSat52-30	30	10,2	0,000394634	46,7
LfrSat53-31	31	9,56	0,000388416	45,2
LfrSat54-29	29	11,32	0,000375723	44,8
LfrSat55-44	44	4,61	0,000351277	70,5
LfrSat56-246	246	11,52	0,000340297	60,2
LfrSat57-27	27	7,72	0,000331941	48,1
LfrSat58-36	36	10,24	0,000327347	47,2
LfrSat59-31	31	11,67	0,000324119	61,3
LfrSat60-30	30	6,8	0,000320317	60
LfrSat61-41	41	14,12	0,000320287	61
LfrSat62-408	408	8,08	0,000294663	57,8
LfrSat63-67	67	5,41	0,000285535	70,1
LfrSat64-541	541	8,2	0,000267099	53
LfrSat65-51	51	8,65	0,000261455	56,9

LfrSat66-94	94	4,24	0,000259752	64,9
LfrSat67-29	29	9,64	0,000233238	55,2
LfrSat68-42	42	10,66	0,000220941	64,3
LfrSat69-60	60	6,98	0,000213188	73,3
LfrSat70-21	21	12,23	0,000212861	66,7
LfrSat71-177	177	8,82	0,000212119	70,1
LfrSat72-32	32	10,69	0,000204396	56,2
LfrSat73-177	177	11,17	0,000193515	65,5
LfrSat74-34	34	4,5	0,00016301	70,6

956

957 **Tabela Suplementar 3:** As características principais dos satélites de *P. semifasciata*.

Satélite	RUL	Divergência	Abundância	A+T(%)
PseSat01-304	304	8,98	0,005458111	68,1
PseSat02-68	68	2,2	0,005339352	66,2
PseSat03-45	45	4,48	0,005313551	62,2
PseSat04-226	226	9,34	0,005006108	64,6
PseSat05-45	45	4,05	0,004210343	44,4
PseSat06-198	198	8,51	0,002821314	66,7
PseSat07-174	174	4,2	0,002811076	62,6
PseSat08-2005	2005	9,46	0,002293053	62,3
PseSat09-50	50	7,42	0,001993153	60
PseSat10-610	610	16,26	0,001795858	66,1
PseSat11-2510	2510	10,79	0,001752841	62,1
PseSat12-2219	2219	3,19	0,001722438	63,9
PseSat13-35	35	1,54	0,001519366	57,1
PseSat14-1235	1235	10,11	0,001139894	57,6
PseSat15-38	38	13,16	0,001078444	63,2
PseSat16-42	42	6,74	0,000959134	59,5
PseSat17-381	381	8,4	0,000843359	61,4
PseSat18-615	615	5,62	0,000710613	58,2
PseSat19-42	42	3,64	0,000687954	64,3
PseSat20-461	461	10,5	0,000665331	63,6
PseSat21-54	54	5,24	0,000662481	57,4
PseSat22-84	84	13,86	0,000648017	53,6
PseSat23-155	155	5,77	0,000647372	63,9
PseSat24-52	52	13,26	0,000617724	55,8
PseSat25-828	828	6,85	0,00058116	51
PseSat26-14	14	6,38	0,00055358	78,6
PseSat27-422	422	6,36	0,000546729	57,6
PseSat28-1284	1284	17,93	0,000524623	56,4
PseSat29-192	192	12,26	0,000451985	64,1
PseSat30-51	51	8,17	0,000429527	39,2
PseSat31-186	186	7,71	0,000396126	71

PseSat32-880	880	2,71	0,000336522	66,9
PseSat33-6	6	22,15	0,000328281	50
PseSat34-165	165	6,23	0,00028134	64,2
PseSat35-23	23	9,4	0,000254937	52,2
PseSat36-195	195	3,03	0,000238349	70,8
PseSat37-300	300	4,17	0,00023586	57,7
PseSat38-80	80	14,83	0,000226345	66,2
PseSat39-1041	1041	9,4	0,000216727	62,3
PseSat40-39	39	8,94	0,000204851	64,1
PseSat41-124	124	6,74	0,000193793	56,5
PseSat42-1093	1093	1,04	0,000193535	60
PseSat43-28	28	8,78	0,000188027	57,1
PseSat44-445	445	3,44	0,000184421	57,3
PseSat45-21	21	16,12	0,00018215	47,6
PseSat46-837	837	6,88	0,000175851	62,5
PseSat47-21	21	10,79	0,000173431	66,7
PseSat48-32	32	5,37	0,000168048	53,1
PseSat49-33	33	5,71	0,000157304	60,6
PseSat50-1125	1125	8,52	0,000155022	65,6
PseSat51-713	713	5,04	0,00015334	62,3
PseSat52-159	159	4,46	0,000141223	67,3
PseSat53-23	23	7,6	0,000137959	56,5
PseSat54-673	673	6,97	0,000137498	57,7
PseSat55-87	87	6,15	0,00013515	67,8
PseSat56-162	162	4,63	0,00012941	67,3
PseSat57-43	43	7,35	0,000125633	48,8
PseSat58-372	372	2,52	0,000120691	56,7
PseSat59-46	46	11,99	0,000118046	60,9
PseSat60-944	944	7,03	0,000111412	61
PseSat61-469	469	3,94	0,000106063	61,4
PseSat62-213	213	4,55	0,000103126	61,5
PseSat63-463	463	3,15	9,94E-05	59,2
PseSat64-187	187	5,11	9,77E-05	62
PseSat65-182	182	9,25	9,22E-05	68,1
PseSat66-44	44	3,87	8,79E-05	61,4
PseSat67-198	198	4,59	8,31E-05	60,1
PseSat68-179	179	5,97	7,67E-05	61,5
PseSat69-759	759	4,27	7,53E-05	58,6
PseSat70-182	182	4,91	6,63E-05	53,8
PseSat71-592	592	3,85	6,46E-05	58,6

958

959 **Tabela Suplementar 4:** As características principais dos satélites de *C. modestus*.

SF	Satélite	RUL	Divergência	Abundância	A+T(%)
----	----------	-----	-------------	------------	--------

	CmoSat01-108	108	12,58	0,0132727	52,8
	CmoSat02-192	192	5,64	0,012516083	64,1
1	CmoSat03-72	72	7,32	0,007128718	47,2
	CmoSat04-187	187	6,98	0,003612853	64,2
	CmoSat05-36	36	8,39	0,003062482	58,3
1	CmoSat06-64	64	6,46	0,002778106	53,1
	CmoSat07-772	772	8,41	0,002089989	59,1
	CmoSat08-319	319	4,57	0,001982271	58,3
	CmoSat09-20	20	10,54	0,001743883	55
	CmoSat10-31	31	4,3	0,00163688	51,6
2	CmoSat11-41	41	8,81	0,001495666	62,6
	CmoSat12-158	158	16,43	0,001320825	49,4
	CmoSat13-2520	2520	11,52	0,001287797	62,7
	CmoSat14-43	43	10,13	0,001214282	30,2
	CmoSat15-647	647	18,75	0,00106864	52,9
	CmoSat16-65	65	6,63	0,001063416	66,2
	CmoSat17-33	33	12,62	0,001061306	72,7
	CmoSat18-50	50	7,44	0,001044099	40
	CmoSat19-1076	1076	9,82	0,00103672	41,5
3	CmoSat20-40	40	18,84	0,000997723	55
	CmoSat21-34	34	12,59	0,000884323	58,8
	CmoSat22-30	30	13,99	0,00084585	46,7
	CmoSat23-41	41	9,26	0,000773475	48,8
	CmoSat24-43	43	4,76	0,000708194	53,5
	CmoSat25-166	166	6,68	0,000694471	51,2
	CmoSat26-41	41	3,98	0,000661497	58,5
	CmoSat27-39	39	10,05	0,000601161	53,8
	CmoSat28-36	36	7,02	0,000589204	69,4
	CmoSat29-87	87	6,39	0,00057935	56,3
	CmoSat30-6	6	24,59	0,000567145	50
3	CmoSat31-24	24	21	0,000552167	45,8
	CmoSat32-31	31	11,35	0,000540568	54,8
	CmoSat33-72	72	8,23	0,000460816	62,5
	CmoSat34-100	100	18,05	0,000427326	56
	CmoSat35-30	30	14,74	0,00037163	56,7
	CmoSat36-991	991	7,92	0,000327283	62,1
	CmoSat37-21	21	9,26	0,000323847	71,4
	CmoSat38-53	53	8,42	0,000322934	56,6
	CmoSat39-23	23	13,28	0,000321847	73,9
	CmoSat40-3008	308	6,91	0,000317073	53,2
	CmoSat41-52	52	12,07	0,000310866	67,3
	CmoSat42-20	20	6,42	0,000303304	55
	CmoSat43-18	18	8,01	0,000303023	44,4

	CmoSat44-66	66	10,54	0,00029434	60,6
	CmoSat45-48	48	6,08	0,000292449	50
	CmoSat46-49	49	8,42	0,000277741	59,2
	CmoSat47-61	61	6,45	0,000260311	54,1
	CmoSat48-25	25	10,48	0,000242095	56
	CmoSat49-946	946	2,79	0,000242	65
	CmoSat50-28	28	5,7	0,000234538	57,1
	CmoSat51-52	52	12,96	0,000233339	53,8
	CmoSat52-695	695	3,76	0,000233196	51,5
	CmoSat53-48	48	6,85	0,000215852	50
	CmoSat54-140	140	8,02	0,000215589	47,9
	CmoSat55-42	42	5,99	0,000208317	64,3
	CmoSat56-22	22	5,96	0,000206569	50
	CmoSat57-48	48	5,72	0,000192151	47,9
	CmoSat58-938	938	2,65	0,000189312	51,5
	CmoSat59-47	47	4,73	0,000183889	40,4
	CmoSat60-16	16	8,8	0,000182659	56,2
	CmoSat61-39	39	9,4	0,000179917	51,3
	CmoSat62-1638	1638	7,71	0,000175825	53,2
	CmoSat63-36	36	0,74	0,000174414	41,7
	CmoSat64-20	20	14,68	0,000172058	55
	CmoSat65-853	853	14,04	0,000162207	59,2
	CmoSat66-22	22	9,15	0,000161574	59,1
	CmoSat67-245	245	6,86	0,000161529	51
	CmoSat68-20	20	4,31	0,000155923	65
	CmoSat69-755	755	1,65	0,000154756	57,6
4	CmoSat70-32	32	10,63	0,000151165	65,6
	CmoSat71-38	38	8,05	0,000149269	55,3
	CmoSat72-164	164	6,66	0,000147333	57,9
	CmoSat73-37	37	3,78	0,000142814	45,9
	CmoSat74-664	664	4,37	0,000142781	55
	CmoSat75-48	48	6,44	0,000141385	56,2
	CmoSat76-30	30	6,68	0,000140386	53,3
	CmoSat77-60	60	8,08	0,000136359	56,7
	CmoSat78-860	860	3,16	0,000129971	52,7
	CmoSat79-240	240	10,03	0,00012954	38,3
	CmoSat80-31	31	5,44	0,000128823	41,9
	CmoSat81-71	71	9,03	0,000126681	60,6
	CmoSat82-654	654	5,05	0,000125145	47,9
	CmoSat83-407	407	12,45	0,000122393	58,2
	CmoSat84-68	68	10,86	0,000122389	61,8
	CmoSat85-37	37	4,5	0,000119764	56,8
	CmoSat86-23	23	9,59	0,000114779	56,5

	CmoSat87-30	30	10,7	0,000114127	56,7
	CmoSat88-213	213	5,35	0,00011293	54,5
	CmoSat89-843	843	1,9	0,000109588	52,7
	CmoSat90-748	748	2,78	0,000108638	57,1
	CmoSat91-776	776	9,8	0,000107563	54,5
	CmoSat92-23	23	7,5	0,000100212	52,2
	CmoSat93-56	56	9,32	9,96327E-05	51,8
	CmoSat94-195	195	5,55	9,91455E-05	52,3
	CmoSat95-350	350	8,76	9,59267E-05	38,6
2	CmoSat96-232	232	2,16	9,30287E-05	41,8
4	CmoSat97-44	44	9,98	9,16644E-05	52,3
	CmoSat98-228	228	9,34	8,44812E-05	52,2
	CmoSat99-123	123	7,83	8,21109E-05	55,3
	CmoSat100-172	172	7,8	8,10416E-05	49,4
	CmoSat101-39	39	6,47	8,05297E-05	56,4
	CmoSat102-319	319	6,51	7,83248E-05	52,7
	CmoSat103-46	46	4,98	6,74624E-05	58,7
	CmoSat104-582	582	1,94	5,4901E-05	61,2
	CmoSat105-434	434	1,36	4,94554E-05	52,3

960

961 **Tabela Suplementar 5:** As características principais dos satélites de *A. mexicanus*.

SF	Satélite	RUL	Divergência	Abundância	A+T(%)
	AmeSat01-1633	1633	7,72	0,006940782	63,3
1	AmeSat02-44	44	6,13	0,005565328	72,7
	AmeSat03-738	738	12,78	0,005121691	54,6
	AmeSat04-178	178	3,36	0,003773901	64,5
	AmeSat05-580	580	5,67	0,002513866	64,8
	AmeSat06-36	36	10,77	0,001293193	40,6
	AmeSat07-328	328	15,14	0,001146492	62,5
	AmeSat08-152	152	4,09	0,001064421	66,4
1	AmeSat09-28	28	11,25	0,001025256	71,4
	AmeSat10-236	236	10,24	0,000818207	64
	AmeSat11-107	107	10,77	0,000654989	53,3
	AmeSat12-796	796	4,56	0,000646847	61,7
	AmeSat13-190	190	6,91	0,000575248	63,7
	AmeSat14-177	177	14,99	0,000555644	67,2
	AmeSat15-6	6	15,03	0,000418275	50
	AmeSat16-70	70	9,82	0,000388785	57,1
	AmeSat17-80	80	9,11	0,000339818	71,2
	AmeSat18-51	51	4,92	0,000312291	58,8
	AmeSat19-52	52	13,09	0,000298545	59,6
	AmeSat20-1266	1266	4,42	0,000291915	55,1

2	AmeSat21-52	52	8,72	0,000277704	67,3
	AmeSat22-85	85	11,24	0,000276953	56,5
	AmeSat23-226	226	13,48	0,000270852	63,7
	AmeSat24-743	743	3,71	0,000255853	56,1
	AmeSat25-186	186	3,17	0,000251404	41,4
	AmeSat26-44	44	11,4	0,000226805	50
	AmeSat27-1245	1245	1,49	0,000222678	53,9
	AmeSat28-42	42	7,67	0,000209881	50
	AmeSat29-742	742	9,47	0,000201802	56,2
	AmeSat30-52	52	14,35	0,000190163	67,3
	AmeSat31-51	51	7,6	0,000184736	54,9
3	AmeSat32-167	167	5,57	0,000180013	72,5
	AmeSat33-175	175	9,56	0,000174543	59,4
3	AmeSat34-173	173	11,29	0,000156503	59
	AmeSat35-658	658	1,74	0,000150961	54
	AmeSat36-353	353	10,31	0,000143743	69,1
	AmeSat37-559	559	12,04	0,000142949	65,3
	AmeSat38-581	581	2,54	0,000131995	57,7
	AmeSat39-577	577	3,22	0,000131839	58,1
	AmeSat40-142	142	5,96	0,000119328	76,8
	AmeSat41-641	641	7,71	0,000112826	57,6
2	AmeSat42-33	33	7,31	0,000108696	75,8
	AmeSat43-372	372	4,8	0,000105366	56,2
2	AmeSat44-24	24	10,55	9,60E-05	70,8
	AmeSat45-48	48	5	9,02E-05	58,3

962

963 **Tabela Suplementar 6:** Coeficiente de correlação de Pearson entre os traços de todos os
964 satelitos das 15 espécies de Characiformes estudadas.

Espécie	RUL x Diverg.	RUL x Abund.	RUL x A+T	Diverg. x Abund.	Diverg. x A+T	Abund. x A+T
<i>C. gomesi</i>	-0,25	0,36	-0,13	0,029	-0,11	-0,022
<i>P. semifasciata</i>	0,0072	0,03	0,085	-0,035	-0,14	0,11
<i>T. auritus</i>	0,047	0,059	0,1	-0,078	0,038	0,075
<i>P. paranae</i>	-0,1	-1,30E-05	-0,11	-0,064	-0,027	-0,048
<i>A. mexicanus</i>	-0,36	0,33	-0,15	-0,0074	0,082	0,12
<i>H. malabaricus</i>	-0,31	-0,078	0,028	-0,12	0,2	-0,036
<i>H. gracilis</i>	-0,53	0,027	-0,41	-0,051	0,14	0,15
<i>P. nattereri</i>	-0,34	0,27	-0,12	0,35	0,14	0,13
<i>M. asterias</i>	-0,5	0,13	0,16	-0,027	0,026	0,085
<i>C. macropomum</i>	0,00073	0,064	-0,011	0,072	0,066	-0,058
<i>P. brachypomus</i>	0,012	0,076	-0,29	-0,016	-0,15	-0,034
<i>P. mesopotamicus</i>	-0,35	0,27	-0,014	0,046	-0,19	-0,06

<i>P. lineatus</i>	-0,37	0,047	0,14	-0,036	0,18	0,21
<i>C. modestus</i>	-0,14	-0,037	0,038	0,069	0,054	0,06
<i>L. friderici</i>	-0,17	0,1	-0,11	-0,061	-0,2	-0,065