

PEDRO HENRIQUE DE MIRA RODRIGUES

Caracterização e análises citogenômicas de satelitomas de Serrasalmideos
(Teleostei, Characiformes)

Bauru

2021

PEDRO HENRIQUE DE MIRA RODRIGUES

Caracterização e análises citogenômicas de satelitomas de Serrasalmideos
(Teleostei, Characiformes)

1
2
3
4
5
6
7
8

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

Orientador: Ricardo Utsunomia

Co-Orientador: Fabio Porto Foresti

Bauru

2021

R696c Rodrigues, Pedro Henrique
Caracterização e análises citogenômicas de satelitomas de
Serrasalmideos (Teleostei, Characiformes) / Pedro Henrique
Rodrigues. -- Bauru, 2021
56 p. : tabs., fotos

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências, Bauru
Orientador: Ricardo Utsunomia
Coorientador: Fabio Porto Foresti

1. Citogenômica. 2. DNA satélite. 3. Serrasalmidae. 4. Satelitoma.
I. Título.

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

Essa ficha não pode ser modificada.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE PEDRO HENRIQUE DE MIRA RODRIGUES, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCIÊNCIAS, DA FACULDADE DE CIÊNCIAS E LETRAS - CÂMPUS DE ASSIS.

Aos 11 dias do mês de maio do ano de 2021, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de PEDRO HENRIQUE DE MIRA RODRIGUES, intitulada **Caracterização e análises citogenômicas de satelitomas de Serrasalmideos (Teleostei, Characiformes)**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. RICARDO UTSUNOMIA (Participação Virtual) do(a) UFRRJ/Rio de Janeiro, Prof. Dr. DIOGO TERUO HASHIMOTO (Participação Virtual) do(a) UNESP/Jaboticabal, Prof. Dr. FRANCISCO JESÚS RUIZ-RUANO CAMPAÑA (Participação Virtual) do(a) University of East Anglia/UK. 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

Quero agradecer a Universidade Estadual Paulista Júlio de Mesquita Filho – UNESP, Bauru e ao departamento de Ciências Biológicas pela formação e aprendizado ao longo de todo trajeto

O Programa de Pós-Graduação em Biociências - Interunidades- pela oportunidade e recursos cedidos.

Gostaria de agradecer aos meus pais, por me darem condições de ingressar em uma Universidade pública de tamanha qualidade como a UNESP e sempre me apoiarem nas minhas escolhas, a minha mãe Maria Teresa, por estar sempre ao meu lado durante esse período, ao meu pai Valdir, que sempre me auxiliou quando necessário.

A minha avó Celia, e ao meu avô Júlio, por todo amor, e todas aquelas velas acessas para me desejar sorte durante um momento difícil.

Em especial a minha avó Ilda, que infelizmente a pouco nos deixou, mas certamente estaria muito orgulhosa de mais uma conquista de um de seus netos.

Ao professor Claudio Oliveira e ao Instituto de Biociências de Botucatu, por todo auxílio na realização das análises desse trabalho

Ao professor Diogo Teruo Hashimoto e ao Centro de Aquicultura da UNESP (CAUNESP) por todo o apoio oferecido para a construção desse trabalho.

A todo pessoal do Lagenpe, pelos ensinamentos e momentos de descontração durante a hora do café ao longo de três anos, todos vocês foram indispensáveis.

Ao professor Fábio Porto-Foresti, por permitir que eu fizesse parte de um lugar tão incrível como o Laboratório de genética de peixes

Ao professor Ricardo Utsunomia, por toda amizade, orientação e conselhos nesses anos de graduação e pós-graduação, certamente um dos maiores exemplos de pesquisador que eu poderia ter tido durante minha passagem acadêmica.

Resumo

Os DNAs satélites (satDNAs) constituem arranjos de sequências repetidas *in tandem*, variando de dezenas até milhares de repetições, que constituem a maior parte da heterocromatina, sendo geralmente encontrados nas regiões pericentroméricas e subteloméricas. Do ponto de vista citogenético, essas sequências são consideradas importantes marcadores cromossômicos, apresentando alta dinâmica evolutiva, tanto por sua característica saltatória nos genomas, quanto pela rápida diversificação de nucleotídeos, resultando na ampla existência de satDNAs espécie-específicos. Os Serrassalmídeos constituem uma família dentro de Characiformes e são conhecidos popularmente como peixes redondos. Animais desse grupo possuem grande importância econômica, pois fazem parte do segundo grupo mais produzido nas pisciculturas. Entretanto, os dados citogenéticos e genômicos desse grupo são escassos e apenas os números diplóides e a distribuição dos genes ribossômicos são conhecidos. O desenvolvimento de plataformas de Sequenciamento de Próxima Geração (NGS) junto de ferramentas bioinformáticas possibilitaram um estudo muito mais simples e eficiente sobre DNAs satélites. Com isso, o objetivo do estudo foi utilizar-se dessas ferramentas para investigar e compreender a evolução dessas sequências repetitivas dentro do genoma de indivíduos da família dos Serrasalmídeos, a partir da caracterização e análise do satelitomas de diferentes espécies do grupo. Para isso, foi utilizada a pipeline Satminer juntamente com as ferramentas RepeatExplorer e TAREAN para a montagem dos satelitomas, seguidas de análises para identificar a diferenciação genética, dinâmica e evolução dessas sequências entre indivíduos do grupo, além de testar novos métodos de identificação de DNAs satélites. Os nossos resultados mostraram um alto nível de compartilhamento entre todas as espécies, seguindo os padrões propostos pela “hipótese *library*” além de apontar um déficit de informação gerado pelos métodos clássicos de caracterização de DNAs satélites

Abstract

Satellite DNAs (satDNAs) constitute tandemly repeated sequences arrays, ranging from a few to hundreds of thousands of repetitions, constituting most of the heterochromatin and generally found in the pericentromeric and subtelomeric regions. From the cytogenetic point of view, these sequences are considered important chromosomal markers, presenting a high evolutionary dynamic, both for their jumping characteristic in the genomes and for the fast nucleotide diversification, resulting in the wide existence of species-specific satDNAs. Serrasalminids, known as round fish, are of economic importance, being part of the second most produced group in fish farms, even so, cytogenetic and genomic data of this group are scarce and only the diploid chromosome number and the distribution of ribosomal genes are known. The development of Next Generation Sequencing (NGS) platforms, with bioinformatics tools made possible a much simpler and more efficient study on satellite DNAs. Thus, the objective of the study was to use these tools to investigate and understand the evolution of these repetitive sequences within the genome of individuals from the Serrasalminids, based on the characterization and analysis of satellitomas of different species in the group. For this, the Satminer pipeline was used combined with the Repeatexplorer and TAREAN tools for the assembly of the satellitomas, followed by analyzes to identify the genetic differentiation, dynamics and evolution of these sequences between individuals in the group, besides to testing new methods of identifying satellite DNAs. Our results showed a high level of sharing among all species, following the patterns proposed by the “library hypothesis” we also point out a deficit of information generated by the classic methods of characterization of DNAs satellites.

Sumário

104		
105	1.1 – Panorama geral sobre DNAs repetitivos, com ênfase em DNAs satélites	6
106	1.2 Histórico e inovação na caracterização de DNAs satélites.....	7
107	1.3 Peixes Characiformes, com foco na família Serralsamidae.....	8
108	2. Objetivos	11
109	2.1 Objetivo geral	11
110	3 Materiais e métodos	12
111	3.1 Amostragem das espécies de Serralsamidae	12
112	3.2 Análises citogenéticas	12
113	3.2.1 Estimulação de mitoses	12
114	3.2.2 Obtenção de cromossomos metafásicos <i>in vivo</i>	13
115	3.3 Coloração convencional com Giemsa	14
116	3.4 Extração de DNA genômico	14
117	3.5 Sequenciamento de nova geração (NGS).....	14
118	3.6 Análises bioinformáticas em servidor remoto	14
119	3.6.1 Identificação de DNAs satélites com Tarean – RepeatExplorer – satMiner	15
120	3.6.2 Análises comparativas.....	15
121	3.7 Genoma mitocondrial – montagem e análises filogenéticas	17
122	3.8 Desenho de <i>primers</i> e preparação das sondas:.....	17
123	3.9 Hibridação <i>in situ</i> fluorescente – FISH	18
124	4 Capítulo 1: Satelitômica comparativa em peixes da família Serralsamidae.....	21
125	4.1 Introdução.....	21
126	4.2 Material e métodos	22
127	4.2.1 Amostras	22
128	4.2.2 Obtenção de suspensões celulares.....	23
129	4.2.3 Sequenciamento do genoma e caracterização do satelitoma	23
130	4.2.4. Análise comparativa entre satelitomas.....	24
131	4.2.8 <i>Minimum spanning tree</i>	25
132	4.2.10 Genoma mitocondrial	25
133	4.2.11 Hibridação <i>in situ</i> fluorescente (FISH).....	25
134	4.3 Resultados	26
135	4.4 Discussão	32
136	5. Referências.....	36
137	6. Material suplementar.....	43
138	7. Considerações finais.....	55

1. Introdução

1.1 – Panorama geral sobre DNAs repetitivos, com ênfase em DNAs satélites

Os genomas dos organismos eucarióticos apresentam uma organização em dois grandes grupos de sequências: o primeiro, composto pelas sequências que apresentam uma única cópia, que compreendem uma porção considerável dos genes funcionais, e o segundo grupo representado pelas sequências de DNA repetitivo, constituído por segmentos de diferentes tipos e tamanhos (Charlesworth et al. 1994). Ademais, as sequências de DNA repetitivo também podem ser classificadas em dois tipos: aquelas dispersas pelo genoma, como os elementos transponíveis; e aquelas que estão repetidas *in tandem*, ou seja, unidades que se repetem uma após a outra, como os DNAs satélites e algumas famílias multigênicas (Charlesworth et al. 1994; Nei e Rooney 2005).

Os DNAs satélites representam uma classe de sequências genômicas altamente repetidas *in tandem* que podem formar longos *arrays* que chegam a milhões de repetições (Miga et al. 2020). Esses elementos genômicos são formados por unidades arranjadas *in tandem*, denominadas monômeros, que podem variar de poucas dezenas até milhares de nucleotídeos (Plohl et al. 2012; Vondrak et al. 2020). Estas sequências estão localizadas principalmente nas regiões pericentroméricas e subteloméricas dos cromossomos, mas também podem ocorrer em regiões intersticiais (López-Flores e Garrido-Ramos 2012). Por representarem uma fração não codificante do genoma, por muito tempo esses fragmentos foram considerados “DNA lixo” (Ruiz-Ruano et al. 2016). Entretanto, estudos recentes têm demonstrado cada vez mais que os DNAs satélites podem desempenhar papéis funcionais dentro das células, como funções centroméricas (Plohl et al. 2014; Escudeiro et al. 2019; Vozdova et al. 2020) e formação da heterocromatina (Lee et al. 2006; Traynor et al. 2019; Otake et al. 2020).

Em geral, as sequências de DNAs satélites evoluem por um processo denominado evolução em concerto, que pode ser dividido em duas etapas: homogeneização e fixação (Lorite et al. 2017). Nesse contexto, a etapa de homogeneização é guiada por mecanismos moleculares como *crossing over* desigual e conversão gênica (Smith 1976; Dover 1986), enquanto a fixação de variantes ocorre por meio de reprodução. Este processo resulta em sequências repetitivas com maior similaridade dentro da espécie do que entre espécies,

ou seja, as múltiplas cópias de uma determinada sequência são homogeneizadas dentro de determinada população (Liao 1999).

Recentemente, Ruiz-Ruano et al. (2016) propuseram uma teoria para a origem e disseminação de DNAs satélites nos genomas de espécies eucarióticas. Em suma, os DNAs satélites surgiriam através de uma duplicação de uma determinada sequência, dando origem a um pequeno *array* de sequências. Posteriormente, esses pequenos *arrays* seriam propagados pelo genoma através de mecanismos como replicação e reinserção, podendo passar por processos de amplificação local. Nesse contexto, à medida que as espécies/populações vão se diferenciando, o conjunto de DNAs satélites nestas amostras vai passando por mudanças majoritariamente quantitativas (Fry e Salser 1977). Por estas razões, de acordo com a “hipótese *library*”, espécies próximas entre si usualmente compartilham parcial ou totalmente o catálogo de DNAs satélites, mas a abundância de cada uma das famílias de DNAs satélites pode variar drasticamente (Utsunomia et al. 2017; Rodrigues et al. 2019; Palacios-Gimenez et al. 2020a). Em casos raros, DNAs satélites podem ser mantidos conservados entre grupos não tão próximos, podendo em casos extremos, apresentar-se em toda uma ordem, conservados ao longo de milhões de anos (Plohl et al. 2010; Petraccioli et al. 2015; Halbach et al. 2020; dos Santos et al. 2021).

1.2 Histórico e inovação na caracterização de DNAs satélites.

O nome "DNA satélite" é histórico, uma vez que este tipo de DNA repetitivo foi descoberto como um pequeno pico no perfil de ultracentrifugação de CsCl (Kit 1961). Embora esta técnica não seja mais utilizada para a caracterização de DNAs satélites, outras metodologias foram extensivamente utilizadas nas últimas décadas, como a cinética de renaturação (Hutton et al. 1973; Botchan, 1974) e digestão com enzimas de restrição (Horz e Zachau, 1977; Kopecka et al. 1978). Apesar de representarem metodologias precisas na caracterização de DNAs satélites, deve-se destacar que a eficiência de cada uma delas não era muito alta, pois representavam metodologias que se baseavam no acaso de selecionar uma determinada enzima que cortasse uma região do monômero de algum DNA satélite. Além disso, por representarem segmentos genômicos altamente dinâmicos e suscetíveis a rápidas mudanças, estes elementos satélites são geralmente sequências espécie- ou gênero-específicas, o que dificulta ainda mais o estudo sistematizado de um mesmo DNA satélite, ou de uma coleção inteira deles (satelitoma) em diferentes espécies (Vicari et al. 2010; Garrido-Ramos 2015; Ruiz-Ruano et al. 2016).

Nesse sentido, a viabilização de metodologias alternativas que permitissem o isolamento, caracterização e análise de sequências repetitivas distribuídas por todo o genoma a custos reduzidos foram extremamente relevantes.

O surgimento e a popularização das tecnologias de sequenciamento de nova geração (NGS, do inglês *next generation sequencing*) permitiram o sequenciamento de genomas inteiros a custos reduzidos, gerando quantidades massivas de dados (Park e Kim 2016). Assim, a combinação destas tecnologias com ferramentas bioinformáticas, permitiu uma alta eficiência na detecção e caracterização de coleções inteiras de DNAs satélites de qualquer organismo não-modelo. Nesse sentido, podemos destacar a abordagem inovadora para a caracterização *de novo* de sequências repetitivas utilizando os *reads* provindos diretamente das plataformas de sequenciamento de DNA descrita por Novák et al. (2013). Destaca-se que até aquele momento, estudos globais com sequências repetitivas eram restritos às espécies com genomas montados, impossibilitando a caracterização destas regiões em organismos não-modelo. No referido estudo, a abordagem consistiu em demonstrar de forma gráfica (*graph-based*), grupos de *reads* que se sobrepõem frequentemente formando *clusters*. Esta estratégia já havia sido utilizada e comprovada por diversos estudos envolvendo plantas e animais (Novák et al. 2010; Macas et al. 2011), mas apenas em 2013, um servidor público implementado na plataforma Galaxy permitiu a popularização deste método.

A partir daquele momento, diversos estudos foram realizados na tentativa de caracterizar catálogos completos de DNAs satélites de diferentes espécies, como no caso do gafanhoto *Locusta migratoria*, no qual foram identificadas 62 famílias de DNA satélite (Ruiz-Ruano et al. 2016) e nos peixes *Psalidodon paranae*, *Characidium gomesi* e *Megaleporinus macrocephalus*, nos quais foram identificados 45, 59 e 164 famílias de DNAs satélites, respectivamente (Silva et al. 2017; Utsunomia et al. 2019; Serrano-Freitas et al. 2020). As constatações aqui apresentadas revelam a diminuição de barreiras metodológicas para o estudo sistematizado de sequências satélites em diferentes espécies.

1.3 Peixes Characiformes, com foco na família Serralsamidae

A região Neotropical se estende desde parte do México, Caribe, até o Sul da América do Sul e abrange cerca de 9.100 espécies de peixes sendo a região com maior diversidade de peixes do mundo, com cerca de 5.160 espécies de peixes de água doce, além de apresentar 27% de espécies de peixes marinhos do mundo (Reis et al. 2016). A

ordem dos Characiformes é uma das maiores ordens dentre os peixes Neotropicais, com aproximadamente 2.200 espécies, divididas em 24 famílias distintas (Eschmeyer e Fong 2020). Dentre elas, a família Serrasalminae é composta pelos popularmente conhecidos “peixes redondos”, sendo subdividida em três diferentes clados: Serrasalminae, composto pelas piranhas; Myleinae, composto pelos pequenos pacus; e Colossomatinae, composto pelos grandes pacus (Kolmann et al. 2020), totalizando 16 gêneros e aproximadamente 101 espécies válidas, amplamente distribuídas em distintas bacias hidrográficas na América do Sul (Ortí et al. 2008). Popularmente conhecido como o grupo dos pacus e piranhas, este clado teve o início de sua diversificação há aproximadamente 60 milhões de anos no Paleoceno médio (Kolmann et al. 2020).

Os representantes da família Serrasalminae apresentam grande importância ecológica, como por exemplo, os pacus que são grandes frugívoros que podem dispersar sementes a longas distâncias (Kolmann, 2020), e econômica, pois são espécies amplamente cultivadas na aquicultura, como os pacus, tambaquis, pirapitingas e seus híbridos, sendo as espécies nativas mais produzidas no Brasil, alcançando valores de centenas de milhares toneladas por ano (IBGE).

De forma geral, poucos estudos citogenéticos foram realizados em peixes Serrasalmídeos, sendo conhecidos basicamente os números diploides de algumas espécies, localização de heterocromatina constitutiva e dos sítios de DNAr 5S e 18S (Gaviria et al. 2005; Baroni et al. 2009; Nakayama et al. 2012; Ribeiro et al. 2014; Ferreira et al. 2016; Favarato et al. 2019). Nesse contexto, algumas espécies apresentam uma estrutura cariotípica com $2n=54$ cromossomos, como as espécies *Colossoma macropomum* e *Piaractus mesopotamicus*, enquanto outras apresentam números diploides mais elevados, como as espécies *Pygocentrus nattereri* e *Serrasalmus elongatus* que apresentam $2n=60$ (Ribeiro et al. 2014) e *Metynnis maculatus* que apresentam $2n=62$ (Baroni et al. 2009; Favarato et al. 2019).

Os dados citogenéticos disponíveis até o momento para a família Serrasalminae reforçam a existência de uma interessante variação cariotípica neste grupo. No entanto, deve-se destacar que algumas porções do genoma dos representantes deste gênero ainda permanecem desconhecidas, como as sequências de DNA repetitivo. Nesse sentido, através da utilização do sequenciamento de nova geração e a consequente caracterização de regiões satélites, espera-se obter informações de interesse para a melhor compreensão da organização do genoma e estrutura cariotípica de quatro espécies relacionadas, aumentando também o conhecimento sobre a diversificação e evolução dessas espécies.

Adicionalmente, deve-se destacar que a espécie *P. nattereri*, conhecida popularmente como piranha-da-barriga-vermelha (*red-bellied piranha*), possui o seu genoma sequenciado e montado (*Pygocentrus nattereri* 1.0.2 assembly, NCBI); nesse sentido, a caracterização da coleção inteira de DNAs satélites desta espécie, também pode permitir uma melhor montagem e compreensão do genoma da piranha-da-barriga-vermelha.

2. Objetivos

2.1 Objetivo geral

Os objetivos gerais do presente trabalho foram:

- 1- Caracterizar os catálogos de DNAs satélites de cinco espécies pertencentes à família Serrasalminidae (*Colossoma macropomum*, *Myleus asterias*, *Piaractus brachypomus*, *Piaractus mesopotamicus* e *Pygocentrus nattereri*) e uma espécie do grupo-irmão Hemiodontidae (*Hemiodus gracilis*).
- 2- Caracterizar a localização cromossômica de DNAs satélites em *Piaractus mesopotamicus*;
- 3- Testar se a hipótese *library* se aplica à evolução dos DNAs satélites na família Serrasalminidae.

3 Materiais e métodos

3.1 Amostragem das espécies de Serrasalminidae

As análises citogenéticas e moleculares na espécie de *P. mesopotamicus* foram realizadas em indivíduos coletados nos tanques de cultivo do Centro de Aquicultura da UNESP, Jaboticabal, São Paulo. Os procedimentos de amostragem, manutenção e análise dos animais foram realizados de acordo com as regras internacionais em experimentação animal seguidas pelo Comitê de Ética o Uso de Animais da Universidade Estadual Paulista (IBB/UNESP), protocolo 1204-CEUA/2019.

Os exemplares analisados foram anestesiados e dissecados, sendo retirados fragmentos do rim anterior para estudos citogenéticos e fragmentos de nadadeira e fígado para posterior extração de DNA e análises moleculares. Os exemplares utilizados na amostragem foram fixados em formol 10%, conservados em etanol 70% e depositados na coleção de peixes do Laboratório de Genética de Peixes, Faculdade de ciências, UNESP-Bauru-SP, sob os números de *voucher* LG13046 até LG13053 para *P. mesopotamicus*.

Além disso, exemplares das espécies *Piaractus brachypomus* e *Colossoma macropomum* também foram amostrados para análises moleculares. Em ambos os casos, fragmentos de nadadeiras foram retirados e fixados em etanol 100% para posterior extração de DNA.

3.2 Análises citogenéticas

Para a análise citogenética dos exemplares utilizamos o método de estimulação de mitoses (Oliveira et al. 1988) e preparação direta de células renais *in vivo* (Foresti et al. 1981) em todos os exemplares analisados, conforme detalhado abaixo.

3.2.1 Estimulação de mitoses

Para a obtenção das preparações cromossômicas, adotamos a técnica de estimulação de divisão celular, que consiste na inserção de uma solução de fermento biológico na musculatura dorsal do animal, como descrito em Oliveira et al. (1988) para peixes. O procedimento utilizado consiste em:

- 1) preparar uma solução de fermento biológico (Fleischmann) na seguinte proporção: 0,5g de fermento, 0,5g de açúcar e 7mL de água destilada;
- 2) incubar a solução em uma estufa (37° C) por cerca de 30 min;

- 3) injetar a solução na região dorsolateral do peixe, na proporção de 1mL para cada 100g de peso do animal. A dose administrada pode ser realizada em duas doses, dividida e aplicada em um período de 48h;
- 4) manter o animal em aquário bem aerado.

3.2.2 Obtenção de cromossomos metafásicos *in vivo*

Para a obtenção de cromossomos metafásicos *in vivo*, seguimos o protocolo descrito por Foresti et al. (1981), utilizado com algumas adaptações. Essa técnica consiste na inibição da polimerização dos microtúbulos através da ação da colchicina, seguida consequentemente de uma hipotonização das células em suspensão, seguindo-se a fixação celular através de uma mistura de metanol/ácido acético (3:1). O procedimento consiste em:

- 1) injetar, na região intra-abdominal, solução aquosa de colchicina (0,025%) na proporção de aproximadamente 1mL para cada 100g de peso do animal;
- 2) deixar o peixe em aquário bem aerado, por um período aproximado de 50 min;
- 3) sacrificar o animal e retirar fragmentos de rim anterior. Transferir o material para uma pequena placa de vidro, contendo 7mL de uma solução hipotônica de KCl (0,075M);
- 4) dissociar o material com o auxílio de pinças de dissecação, complementando esse processo com o auxílio de uma pipeta seringa de 3mL, até que se obtenha uma solução aquosa homogênea;
- 5) transferir a solução obtida para um tubo de centrífuga e incubar esta solução em estufa a 37°C por 21 min;
- 6) retirar o tubo da estufa, colocando 7 gotas de fixador gelado (metanol e ácido acético na proporção de 3:1, respectivamente); agitar levemente a mistura com uma pipeta Pasteur e deixar repousar por 5 min à temperatura ambiente;
- 7) adicionar cerca de 6mL de fixador e novamente agitar a mistura; levar à centrífuga (900 ± 100 rpm) por 10 min;
- 8) retirar o sobrenadante, e logo após ressuspender o material precipitado em 7mL de fixador; centrifugar por 7 min a 900 ± 100 rpm;
- 9) repetir o item 8 por três vezes para uma completa fixação das células em suspensão;
- 10) pingar o material em lâminas previamente tratadas em solução de HCL 10%;
- 11) deixar secar ao ar.

3.3 Coloração convencional com Giemsa

As preparações cromossômicas fixadas nas lâminas foram coradas por dez minutos com solução de Giemsa a 5 % em tampão fosfato (pH=6,8) para a verificação do número diploide das espécies *P. mesopotamicus* e *C. macropomum*.

3.4 Extração de DNA genômico

O DNA genômico foi obtido a partir de amostras de fígado ou nadadeira, preservados em etanol 100%, utilizando os Kits Wizard Genomic DNA Purification (PROMEGA) ou DNeasy Blood & Tissue (QIAGEN), conforme instruções recomendadas pelo fabricante, com um passo para a remoção de RNA com RNase A (Invitrogen). Os produtos gerados foram visualizados em eletroforese de gel de agarose a 1%.

3.5 Sequenciamento de nova geração (NGS)

Para o presente estudo, foram utilizadas bibliotecas genômicas de 5 espécies pertencentes à família Serrasalmidae e uma espécie pertencente à família Hemiodontidae, grupo-irmão de Serrasalmidae. O sequenciamento massivo de DNA pela metodologia de nova geração foi realizado na plataforma Illumina MiSeq (*paired-end* 2x251pb em *Piaractus mesopotamicus*, *Myleus asterias* e *Hemiodus gracilis*) e na plataforma MGISEQ-2000 (*paired-end* 2x101 em *Colossoma macropomum* e *Piaractus brachypomus*). Além disso, uma biblioteca genômica Illumina HiSeq2500 de *Pygocentrus nattereri* (*paired-end* 2x250pb; SRA Accession number: SRR3136983) também foi utilizada.

3.6 Análises bioinformáticas em servidor remoto

As análises bioinformáticas foram realizadas em servidor remoto da Universidade Estadual Paulista, que está localizado no Instituto de Biociências de Botucatu da Universidade Estadual Paulista, sob os cuidados do Dr. Claudio de Oliveira. Um mesmo *workflow* foi utilizado para cada uma das bibliotecas sequenciadas, utilizando *scripts* descritos em Ruiz Ruano et al. (2016) e disponíveis no endereço (<https://github.com/fjruizruano/ngs-protocols>).

3.6.1 Identificação de DNAs satélites com Tarean – RepeatExplorer – satMiner

A caracterização dos satelitomas das 6 espécies foram realizadas utilizando as ferramentas TAREAN (Novák et al. 2017) e satMiner (Ruiz-Ruano et al. 2016) em conjunto, seguindo os seguintes passos: i) preparar os *reads* para aplicação no programa TAREAN (Novak et al. 2017), já com os cortes nos adaptadores, seleção de qualidade e *reads* cortados em 100bp; ii) rodar o *software* TAREAN com a opção padrão; iii) isolar as sequências consenso com maior *k-mer coverage score* dos clusters apontados como *high* e *low confidence putative satellites*; iv) montar uma base de dados customizada com os satélites já isolados e utilizar um *script* (deconseq_run.py) (Schmieder and Edwards, 2011), para exclusão de *reads* semelhantes aos satélites já identificados da biblioteca inicial; v) selecionar novos *reads* para uma nova rodada de TAREAN; vi) realizar as etapas (i a v) até que nenhum DNA satélite novo seja encontrado; vii) utilizar as bibliotecas descontaminadas para preparar *reads* para aplicação no programa RepeatExplorer1; viii) rodar o *software* Repeatexplorer1 com a opção padrão; ix) selecionar manualmente, com auxílio do *software* Geneious 8.4, os clusters que apresentarem a forma de gráficos circulares (formato padrão de DNAs satélites); x) repetir as etapas iv e v; xi) repetir as etapas (vii a x) até não ser mais encontrado sequências repetidas *in tandem*.

Após a identificação de todos os satélites, um *script* (rm_homology.py) para identificação de satélites similares foi aplicado, adotando como critério de homologia >95% de similaridade como mesmo satélite, >80% como variante e >50% como superfamília (Ruiz-Ruano et al. 2016). Com uma biblioteca final para cada espécie, o *software* RepeatMasker v4.0.5 (Smit et al. 2017) foi utilizado para classificar os DNAs satélites encontrados de acordo com a sua abundância e verificar a variação dessas sequências no genoma.

3.6.2 Análises comparativas

Após a caracterização dos seis satelitomas de forma independente, quatro diferentes estratégias de comparação foram aplicadas, uma vez que nenhuma análise comparativa extensiva entre satelitomas foi realizada entre múltiplas espécies até o momento. A seguir, detalhamos cada uma das abordagens utilizadas:

- 1- A primeira abordagem consistiu em uma comparação entre todos os catálogos isolados brutos imediatamente após os protocolos de TAREAN e satMiner

utilizando o *script* rm_homology.py, seguido de alinhamentos no algoritmo MUSCLE (Edgar 2004) entre as sequências similares. A partir disso, um catálogo único contendo todos os satélites de todas as espécies que apresentaram menos de 80% de similaridade foi construído. Esse catálogo foi comparado com os satelitomas de peixes Neotropicais disponíveis na literatura. Além disso, todos os satDNAs com mais de 80% de similaridade encontrados em diferentes espécies foram identificados como satDNAs compartilhados entre espécies;

- 2- Uma segunda estratégia aplicada foi realizar uma busca forçada pelos satélites ausentes em cada uma das espécies na estratégia anterior, para obter uma tabela definitiva de presença/ausência dessas sequências, uma vez que é sabido que apenas satélites com um mínimo de abundância será localizado pelos protocolos utilizados. Para isso, utilizamos o *software* RepeatMasker v4.0.5 (Smit et al. 2017) em cada uma das espécies para buscar os satélites que não haviam sido encontrados durante a caracterização do satelitoma. De forma mais específica, fizemos uma busca direcionada, utilizando a biblioteca total de cada espécie, dos *reads* que apresentaram similaridade com os satDNAs-alvo, utilizando uma coleção de *scripts* personalizados gentilmente cedidos pelo Dr. Francisco J. Ruiz Ruano e submetemos estes *reads* ao TAREAN para clusterização utilizando uma *custom database* com os satélites-alvo. A partir disso, as sequências consenso dos clusters apresentados como *high* e *low confidence putative satélites* foram analisadas detalhadamente para verificar se o satDNA procurado estava na lista de sequências consenso com mais de 80% de similaridade entre os monômeros. Por vezes, alguns DNAs satélites não eram classificados como DNA satélite putativo, mas a análise dos clusters do RepeatExplorer2 através de *self-dotplots* evidenciava repetições *in tandem* em seus *contigs*. Neste caso, considera-se que os clusters referidos, não alcançaram os *scores* necessários para serem classificados como satDNAs pelo TAREAN por apresentarem uma organização um pouco mais complexa;
- 3- Uma terceira abordagem consistiu em utilizar as sequências de todos os DNAs satélites no mínimo dimeradas (até alcançarem 200bp, pelo menos) e mapear os *reads* de cada uma das espécies utilizando o *software* bowtie2 (Langmead e Salzberg 2012) para examinar os perfis de variação do número de cópias

utilizando o programa RepeatProfiler (Negm et al. 2021). As configurações -
sensitive e -*no-mixed* do bowtie2 foram utilizadas para aumentar as chances
de que apenas *reads* provindos de regiões em *tandem* fossem mapeados;

- 4- Finalmente, uma quarta abordagem consistiu em utilizar o catálogo completo
de todas as espécies (construído durante a primeira abordagem, ver acima) e
avaliar os valores de abundância e divergência em cada uma das bibliotecas
utilizando o *software* RepeatMasker v4.0.5 (Smit et al. 2017). Embora esta
análise não indica a ocorrência de organização em *tandem* dos satélites
buscados, é uma análise de alta sensibilidade e que pode indicar a ocorrência
de resquícios de sequência, mesmo que não estejam arranjadas em *tandem*.

Ao final das análises, matrizes de presença/ausência de DNAs satélites foram
construídas separadamente para as abordagens 1 e 2. Finalmente, matrizes de distância
euclidiana para presença de satDNAs em cada espécie foram construídas.

3.7 Genoma mitocondrial – montagem e análises filogenéticas

Os genomas mitocondriais das 6 espécies foram montados com o auxílio do
software NOVOPlasty (Dierckxsens et al. 2017), utilizando as bibliotecas *short-reads* e
a sequência COI de cada espécie como *seed*. Após a montagem, os mitogenomas foram
anotados utilizando o programa MitoFish (Iwasaki et al. 2013) e as sequências foram
alinhas utilizando o algoritmo MAFFT v7.407 (Katoh e Toh 2010), sendo que as
regiões com *gaps* e não-informativas foram removidas do alinhamento final utilizando o
software Gblocks v0.91b (Talavera e Castresana 2007). Posteriormente, calculamos as *p-*
distances entre cada um dos mitogenomas para a verificar a correlação entre presença de
satDNAs e distâncias entre espécies.

3.8 Desenho de *primers* e preparação das sondas:

Com os satélites de *Piaractus mesopotamicus* caracterizados, *primers* foram
desenhados para o preparo de sondas por PCR. Para os satélites com mais de 30 bp,
primers divergentes (Figura 1) foram desenhados, como visto no esquema abaixo. A
qualidade dos *primers* foi verificada com o *Multiple Primer Analyzer* (ThermoFisher
Scientific), de maneira a evitar a formação de grampos e auto-pareamento.

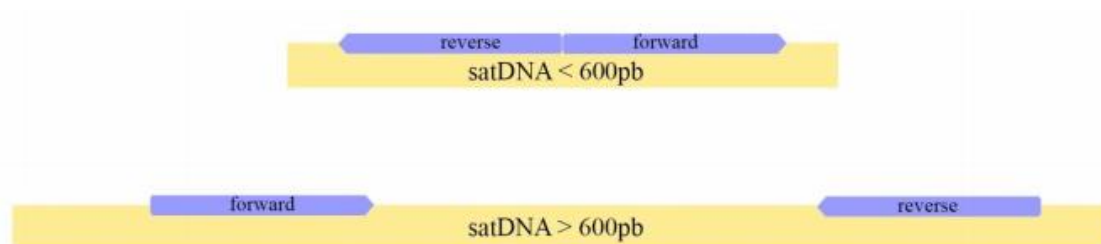


Figura 1. Exemplo dos *primers* divergente e convergente

Reações de Amplificação em Cadeia da Polimerase (PCR) foram realizadas para cada um dos DNAs satélites, a partir do DNA genômico de *P. mesopotamicus* em diferentes diluições seriadas com os seguintes parâmetros: 1x PCR Buffer, 1.5mM de MgCl₂, 200μM de dNTP, 0,1 μM de Primer A + Primer B, 0,1U de Taq polimerase e 2-100ng de DNA genômico, em um total da reação de 12,5 μL. A produção das sondas seguiu o mesmo protocolo, com a incorporação de digoxigenina-11-dUTP (Roche Applied Science).

Devido ao alto número de satélites, as reações em termociclador variaram de acordo com os parâmetros exigidos para cada DNA satélite.

95° C → 6 minutos	}	30x
95° C → 20 segundos		
50°-56° C → 10-50 segundos		
72° C → 10-50 segundos		
72° C → 14 minutos		

3.9 Hibridação *in situ* fluorescente – FISH

As sondas utilizadas na FISH de todos os DNAs satélites deste trabalho foram marcadas com digoxigenina-11-dUTP, tendo como base a amplificação das sequências satélites nos genomas de *P. mesopotamicus*. Para a realização da técnica de FISH utilizamos os procedimentos descritos por Pinkel et al. (1986), utilizando algumas modificações já estabelecidas no Laboratório de Genética de Peixes (UNESP/Bauru):

Primeiro dia:

Primeira lavagem:

1. Lavar as lâminas em solução PBS 1x, por 3 min.

Tratamento com Pepsina:

1. Incubar em 0,005% pepsina/10mM HCl por 10 min, a 37 °C;
2. Lavar em PBS 1x por 3 min, em temperatura ambiente.

Fixação com formaldeído

1. Fixar em 1% em PBS 1x/50 mM MgCl₂ por 10 min, em temperatura ambiente;
2. Lavar em PBS 1x por 3 min;
3. Desidratar em série alcoólica (70%,85% e 100%) gelada por 3 min cada e secar as lâminas.

Mix de Hibridação

1. Preparar o *mix* de hibridação (15 ul de formamida 70%, 6 ul de sulfato dextrano 10%, 3 ul 20xSSC, entre 1 ou 2 ul de sonda; por lâmina, completando de água até o total de 30ul);
2. Ligar banho seco (ou termociclador) a 98°C.

Pré-Hibridação

1. Desnaturar as lâminas em formamida 70% em 2xSSC, por 2 min a 70°C;
2. Desnaturar a sonda no banho seco por 10 min;
3. Desidratar as lâminas em série alcoólica, 3 min cada e secar as lâminas.

Hibridação

1. Montar as lâminas com a solução de hibridação, cobrir com lamínula e incubar em câmara úmida a 37°C overnight.

Segundo dia:

Lavagens:

1. Retirar as lamínulas e lavar as lâminas em 1x SSC em 42°C por 5min, realizando a agitação do recipiente;
2. Lavar as lâminas em 4xSSC/0,5% Triton por 3min, em temperatura ambiente realizando a agitação do recipiente.

Bloqueio

1. Incubar as lâminas em Tampão 5% NFDm/4xSSC por 15 min a temperatura ambiente;

2. Lavar 2x em 4xSSC/0,5% Triton por 3 min cada, a temperatura ambiente;

Detecção

1. Incubar as lâminas com 100 ul de NFDM/0,5 ul de anti-digoxigenina rodamina por 1h em câmara úmida a 37°C;

2. Lavar 3x em 4xSSC/0,5% Triton por 3 min cada, em temperatura ambiente;

3. Desidratar em série alcoólica, 3 min cada e secar as lâminas;

4. Montar as lâminas inserindo 15 ul de DAPI e cobrir com lamínula, mantendo estas em um local escuro até sua visualização.

4 Capítulo 1: Satelitômica comparativa em peixes da família Serrasalminidae

4.1 Introdução

O genoma dos eucariotos é constituído principalmente por sequências repetitivas, que podem estar dispersas pelo genoma, como os elementos transponíveis (TEs), ou repetidas *in tandem*, como os DNAs satélites e as famílias multigênicas (Charlesworth, 1994). Os DNA satélites são sequências não-codificantes, que podem formar longos *arrays*, localizados essencialmente nas regiões pericentroméricas e subteloméricas, formando a maior parte da heterocromatina (Garrido-ramos 2017). De acordo com a *library hypothesis* (Fry e Salser 1977), essas sequências apresentam uma origem comum entre organismos próximos, amplificando-se de forma independente em cada organismo, gerando diferentes perfis de abundância de satélites em espécies próximas.

Em geral, a evolução molecular das sequências de DNAs satélites é guiada por um processo denominado evolução em concerto, que pode ser dividido em duas etapas: homogeneização e fixação (Smith 1976; Dover 1986; Liao 1999). Este processo resulta em sequências repetitivas com maior similaridade dentro da espécie do que entre espécies, ou seja, as múltiplas cópias de uma determinada sequência são homogeneizadas dentro de determinada população (Liao 1999). Nesse contexto, é de se esperar que o compartilhamento de satDNAs homólogos entre espécies próximas seja menor conforme a distância entre as espécies aumenta. Entretanto, esta hipótese foi pouco testada, sendo restrita a sequências satélites curtas (até 40bp) em grupos com maiores quantidades de recursos genômicos (Flynn et al. 2018; Wei et al. 2018; Cechova et al. 2019).

Até o presente momento, poucos estudos a nível de satelitoma foram realizados em peixes Neotropicais, sendo que todos visaram uma melhor caracterização da estrutura de variados cromossomos sexuais e/ou supranumerários, fornecendo indícios da diferenciação destes elementos em diferentes linhagens (Silva et al. 2017; Utsunomia et al. 2019; Serrano-Freitas et al. 2020). Além disso, os catálogos de satDNAs gerados nestes estudos permitiram uma análise comparativa entre três satelitomas (totalizando 268 DNAs satélites), levando à caracterização de um único DNA satélite conservado em toda a ordem dos Characiformes, denominado CharSat01, desafiando as previsões de evolução molecular de satDNAs (dos Santos et al. 2021). No entanto, deve-se destacar que a comparação entre os satelitomas se deu em nível de catálogos caracterizados por meio de RepeatExplorer, ou seja, a não-detecção de outros DNAs satélites compartilhados entre as espécies não indicam, necessariamente, a ausência destas sequências nestas espécies.

Serrasalminae é uma família inserida na ordem dos Characiformes e inclui, atualmente, 16 gêneros e aproximadamente 101 espécies válidas (Eschmeyer e Fong 2020), amplamente distribuídas em distintas bacias hidrográficas na América do Sul (Ortí et al. 2008). Atualmente, este grupo pode ser subdividido em 3 clados principais com tempos de divergência definidos: Colossomatinae, sendo o grupo dos pacus grandes; Mylenae, abrangendo os pacus pequenos; e os Serrasalminae, representando as piranhas (Kolmann et al. 2021). Nesse sentido, a avaliação da dinâmica de satDNAs ao longo deste grupo pode fornecer respostas interessantes sobre a persistência evolutiva destas sequências.

Diante do exposto, o presente estudo teve como objetivo caracterizar e comparar os satelitos de 5 espécies de Serrasalminae (*Colossoma macropomum*, *Mylenasterias*, *Piaractus brachipomus*, *Piaractus mesopotamicus* e *Pygocentrus nattereri*) e 1 de Hemiodontidae, grupo-externo a Serrasalminae (*H. gracilis*), a partir de bibliotecas genômicas produzidas por sequenciamento de segunda geração utilizando diferentes protocolos. Além disso, nós mapeamos várias famílias de satDNAs nos cromossomos de *P. mesopotamicus* para identificar a organização cromossômica destas sequências. Nossos resultados revelaram uma ampla persistência de sequências de DNAs satélites ao longo de 80 milhões de anos. Além disso, demonstramos que a comparação pura e simples de catálogos de DNAs satélites não é suficiente para análises integrativas futuras, uma vez que uma grande porcentagem de sequências repetidas em *tandem* não é caracterizada primariamente pelos métodos tradicionais *de novo* de caracterização de satDNAs.

4.2 Material e métodos

4.2.1 Amostras

Indivíduos de *P. mesopotamicus* foram coletados nos tanques de cultivo do Centro de Aquicultura da UNESP (CAUNESP), Jaboticabal, São Paulo, sob a responsabilidade e colaboração do Dr. Diogo Teruo Hashimoto. Os procedimentos de amostragem, manutenção e análise dos animais foram realizados de acordo com as regras internacionais em experimentação animal seguidas pela Universidade Estadual Paulista, autorização 1204-CEUA/2019. Os exemplares utilizados na amostragem foram fixados em formol 10%, conservados em álcool 70% e depositados na coleção de peixes do Laboratório de Genética de Peixes, Departamento de Ciências Biológicas, UNESP-Bauru-SP. Além disso, exemplares das espécies *Piaractus brachipomus* e *Colossoma*

macropomum também foram amostrados para análises moleculares. Em ambos os casos, fragmentos de nadadeiras de indivíduos mantidos nos tanques de aquicultura do CAUNESP foram retirados e fixados em etanol 100% para posterior extração de DNA.

4.2.2 Obtenção de suspensões celulares

Para a obtenção de cromossomos metafásicos, o protocolo descrito por Foresti et al. (1981) foi utilizado, sendo que os cromossomos mitóticos foram obtidos a partir de suspensões celulares do rim anterior dos indivíduos de *P. mesopotamicus* (8 indivíduos).

4.2.3 Sequenciamento do genoma e caracterização do satelitoma

Para a caracterização *de novo* de DNAs satélites dos genomas de *Piaractus mesopotamicus*, *Myleus asterias* e *Hemiodus gracilis*, foram utilizadas bibliotecas de sequenciamento Illumina MiSeq (*paired-end* 2x251bp), enquanto para *Colossoma macropomum* e *Piaractus brachypomus* foi utilizado o sequenciamento MGISEQ-2000 (*paired-end* 2x101), além de uma biblioteca genômica Illumina HiSeq2500 de *Pygocentrus nattereri* (*paired-end* 2x250pb; SRA Accession number: SRR3136983). Inicialmente, uma etapa de filtro de qualidade e tamanho de *reads* foi aplicada utilizando o *software* Trimmomatic (opções utilizadas: MINLEN:100 CROP:100) (Bolger et al. 2014). Posteriormente, 500,000 *paired-end reads* foram selecionados para serem clusterizados com o programa TAREAN (Novák et al. 2017), com as opções padrão, visando a identificação, de forma automática, de DNAs satélites putativos (*high e low confidence*). Os monômeros identificados foram extraídos da biblioteca original com o uso do *software* Deconseq (Schmieder e Edwards 2011), repetindo o processo até não serem mais identificados DNAs satélites. Uma vez que o programa não retornou novos DNAs satélites putativos, demos início às buscas manuais utilizando o RepeatExplorer1 (Novák et al. 2013) com as opções padrão, buscando clusters de formato circular, assim como descrito em Ruiz-Ruano et al. (2016). Subsequentemente, os *contigs* associados a estes clusters foram analisados com o *software* Geneious 8.1, buscando sequências repetidas *in tandem* através dos gráficos *dotplot*. Os monômeros identificados aqui, também foram extraídos da biblioteca original com o uso do *software* Deconseq (Schmieder e Edwards 2011), sendo que este *workflow* foi repetido até que não fossem mais encontrados DNAs satélites clusterizados.

Após a fase de caracterização, a homologia entre os DNAs satélites caracterizados foi analisada, sendo classificadas como variante (>95%), família (>80%), e superfamília (>50%) (Ruiz-Ruano et al. 2016). Posteriormente, os resultados foram comparados entre si e com os satelitomas de peixes já descritos, utilizando o mesmo critério de classificação. O *software* RepeatMasker (Smit 2017), foi utilizado para determinar a abundância e a divergência de cada variante dentro do genoma.

4.2.4. Análise comparativa entre satelitomas

Após a caracterização dos seis satelitomas de forma independente, quatro estratégias de comparação entre satelitomas foram aplicadas. A seguir, detalhamos cada uma das abordagens utilizadas: i) comparação entre catálogos de DNAs satélites isolados independentemente de cada espécie. Nesta abordagem, detectamos traços de similaridade entre as sequências utilizando o *software* RepeatMasker implementado no *script* personalizado `rm_homology.py` (<https://github.com/fjruizruano/ngs-protocols>) e alinhamos as sequências identificadas com o algoritmo MUSCLE (Edgar 2004), após o alinhamento, os satDNAs que apresentaram mais de 80% de similaridade foram considerados conservados entre espécies. Após estes passos, conseguimos gerar um catálogo geral de satDNAs para as espécies-alvo, que foi utilizado em análises *downstream*; ii) A grande maioria dos satDNAs encontrados na primeira abordagem não estava presente em todas as espécies. Nesse sentido, cada DNA satélite do catálogo geral das espécies aqui analisadas foi submetido a buscas manuais em cada uma das bibliotecas genômicas disponíveis (1,000,000 de *reads paired-end*). Para isso, selecionamos *reads* homólogos de cada DNA satélite ausente nas espécies usando BLAT (Kent 2002) e clusterizamos com a ferramenta TAREAN; iii) Uma terceira abordagem incluiu selecionar todos satélites do catálogo, dimerar os monômeros até, no mínimo, 200nt em um arquivo de referência e aplicar a *pipeline* RepeatProfiler (Negm et al. 2021) em cada biblioteca (10,000,000 de *reads paired-end*) independentemente. Também utilizamos dez genes de peixes de cópia única para serem mapeados para normalização da cobertura de leitura (`ppfia1` [XM_022685633.1], `foxl2` [XM_007232295.3], `prospero` [XM_017708821.1], `msh4` [XM_017711771.1], `zdhhc22` [XM_017711775.1], `coq6` [XM_017711829.1], `znf106` [XM_017711848.1], `lactamase` [XM_022682177.1], `gastrula zinc finger` [XM_022685636.1], and `tubulin-kinase` [XM_017711762.1]); iv) Finalmente, uma quarta abordagem foi utilizar o catálogo geral de satDNAs como

referência e realizar uma busca com o *software* RepeatMasker (Smit 2017) em cada biblioteca (1,000,000 de *reads paired-end*), caso houvesse algum read mapeado, já consideramos a presença de determinada sequência.

4.2.8 *Minimum spanning tree*

Entre os DNAs satélites identificados em todas espécies, três foram selecionados aleatoriamente para criação de uma *Minimum spanning tree* (MST), sendo dois deles clusterizados em todas espécies e um com clusterização exclusiva de uma espécie, porém com sinais RepeatMasker em ambas, para a visualização do comportamento dos monômeros dessas sequências entre os genomas

4.2.10 Genoma mitocondrial

Os genomas mitocôndriais das 6 espécies foram montados com o auxílio do *software* NOVOPlasty (Dierckxsens et al. 2016), utilizando as bibliotecas *short-reads* e a sequência COI de cada espécie como *seed*. Após a montagem, os mitogenomas foram anotados utilizando o programa MitoFish (Iwasaki et al. 2013) e as sequências foram alinhadas utilizando o algoritmo MAFFT v7.407 (Katoh e Toh 2010), sendo que as regiões com *gaps* e não-informativas foram removidas do alinhamento final utilizando o *software* Gblocks v0.91b (Talavera e Castresana 2007). Posteriormente, calculamos as *p-distances* entre cada um dos mitogenomas para verificar a correlação entre presença de satDNAs e distâncias entre espécies.

4.2.11 Hibridação *in situ* fluorescente (FISH)

Primers divergentes (em orientação oposta) foram desenhados para todos os DNAs satélites que apresentaram tamanho maior que 30 bp (Tabela S1). Reações de PCR foram realizadas e, após confirmação de amplificação inicial, reações independentes de PCR foram realizadas para a produção de sondas marcadas com digoxigenina-11-dUTP. Os experimentos de FISH foram realizados de acordo com os procedimentos descritos por Pinkel et al. (1986) e os sinais foram detectados com anti-digoxigenina-rhodamina. Para uma correta avaliação da distribuição cromossômica dos DNAs satélites, foram analisadas aproximadamente 10 metáfases por indivíduo. As hibridações foram realizadas utilizando preparações provenientes do mesmo indivíduo.

4.3 Resultados

Para *P. mesopotamicus*, após duas iterações de TAREAN e duas do protocolo satMiner, nós identificamos 30 famílias de DNAs satélites distintas, com comprimento de unidades repetidas (RUL, do inglês *repeat unit lenght*) variando de 6 bp até 1853 bp (média de 286 bp), apresentando um equilíbrio entre DNAs satélites curtos e longos, onde 16 DNAs satélites apresentaram RUL menor que 100 bp e 14 apresentaram RUL maior que 100 bp. A composição A+T variou de 45.2% até 76.1% entre as famílias, com média de 59.6% indicando um leve viés em relação aos DNAs satélites ricos em A+T (Tabela S1). Dentre as 30 famílias, uma única superfamília foi identificada, sendo composta pelos dois DNAs satélites mais abundantes (PmeSat01 e PmeSat02), com RULs de 508 e 143 respectivamente.

Para *P. brachypomus*, após sete iterações de TAREAN e uma de satMiner, identificamos 63 DNAs satélites, sendo o maior resultado obtido entre os Serrasalmideos, suas RULs variaram de 6 a 4.663 bp (média de 862bp) com dominância de DNAs satélites longos com 47 sequências contra 16 curtas. Sua composição A+T variou de 44,40% até 73,60% com média de 58,23% (Tabela S2). 6 superfamílias foram identificadas, PbrSat01 e 02; PbrSat02 e 27; PbrSat03, 11, 15, e 22; PbrSat 07, 35 e 50; PbrSat08 e 20; PbrSat36 e 37; e PbrSat44 e 49.

Para *C. macropomum*, após duas iterações de TAREAN e três de satMiner, foram identificados 46 DNAs satélites, com RULs variando de 6 a 3168 bp (média de 429bp), com dominância de DNAs satélites longos com 27 sequências contra 19 curtos. A composição A+T variou de 43.3% até 76.2%, com média de 59.3% (Tabela S3). Duas superfamílias foram identificadas, CmaSat01 e CmaSat04; e CmaSat05 e CmaSat32.

Para *M. asterias*, após duas iterações de TAREAN e duas de satMiner obtivemos 31 DNAs satélites, com RULs variando de 6 a 946 bp (média de 113 bp), com dominância de DNAs satélites curtos com 22 sequências, contra 9 longas. Sua composição A+T variou de 36,53% até 76,19% com média de 58,42% (Tabela S4). Nenhuma superfamília foi identificada

Para *P. nattereri*, após duas iterações de TAREAN e duas de satMiner foram identificados 57 DNAs satélites, com RULs variando de 6 a 2469 bp (média de 472bp) com equilíbrio entre DNAs satélites curtos (27) e longos (30), sua composição A+T variou de 44,84% até 76,19% com média de 57,91% (Tabela S5). Foram identificadas 5 superfamílias, sendo elas PnaSat23 e 36; e PnaSat37 e 44.

Para *H. gracilis*, após duas iterações de TAREAN e duas de satMiner foram identificados 59 DNAs satélites, com RULs variando de 6 a 1215 bp (média de 281 bp), com equilíbrio entre DNAs satélites curtos (30) e longos (29). Sua composição A+T variou de 44,47% até 76,19% com média de 55,58% (Tabela S6). Cinco superfamílias foram identificadas, HgrSat01, 04 e 05; HgrSat02 e 08; HgrSat03, 22, 23 e 28; HgrSat12 e 44; e HgrSat02 e 18

Analisando o número de DNAs satélites compartilhados entre os catálogos de satélites brutos (Figura 1a), observamos que as espécies pertencentes ao clado dos Colossomatinae (*C. macropomum*, *P. brachypomus* e *P. mesopotamicus*) apresentaram maior proximidade entre si, sendo que as duas espécies pertencentes ao gênero *Piaractus*, apresentaram o maior número de satDNA compartilhados (21 sequências compartilhadas) (Figura 1a). *M. asterias* e *P. nattereri*, apresentaram resultados equilibrados em relação a todas espécies, enquanto que *H. gracilis*, representante do grupo Hemiodontidae, grupo externo a Serrasalminidae, apresentou os menores números de sequências compartilhadas entre todas as espécies. No total, 34 DNAs satélites estiveram presente em ao menos 2 espécies, enquanto apenas 6 estão presentes em todas as espécies (Tabela suplementar 1).

Utilizando a clusterização forçada dos satDNAs, identificamos diversos DNAs satélites não encontrados pelos métodos convencionais de caracterização de satelitomas. O que pode ser justificado pela menor abundância destas sequências ou mesmo por uma forma diferencial de organização. Nesse sentido, foram encontrados mais 24 DNAs satélites em *P. mesopotamicus*, 11 em *P. brachypomus*, 8 em *C. macropomum*, 7 em *P. nattereri*, 5 em *M. asterias* e 3 em *H. gracilis*. Além disso, utilizando a ferramenta RepeatExplorer2, foram identificados mais 10 DNAs satélites para *P. mesopotamicus*, 7 para *P. brachypomus*, 8 para *C. macropomum*, *M. asteris* e *P. nattereri* e 5 para *H. gracilis*. Após avaliação, obtivemos um resultado final de presença/ausência de DNAs satélites entre as espécies aqui estudadas, com alterações substanciais em relação ao primeiro diagrama de Venn (Figura 1b). Destaca-se o aumento de sequências compartilhadas entre todas as espécies de 6 para 17.

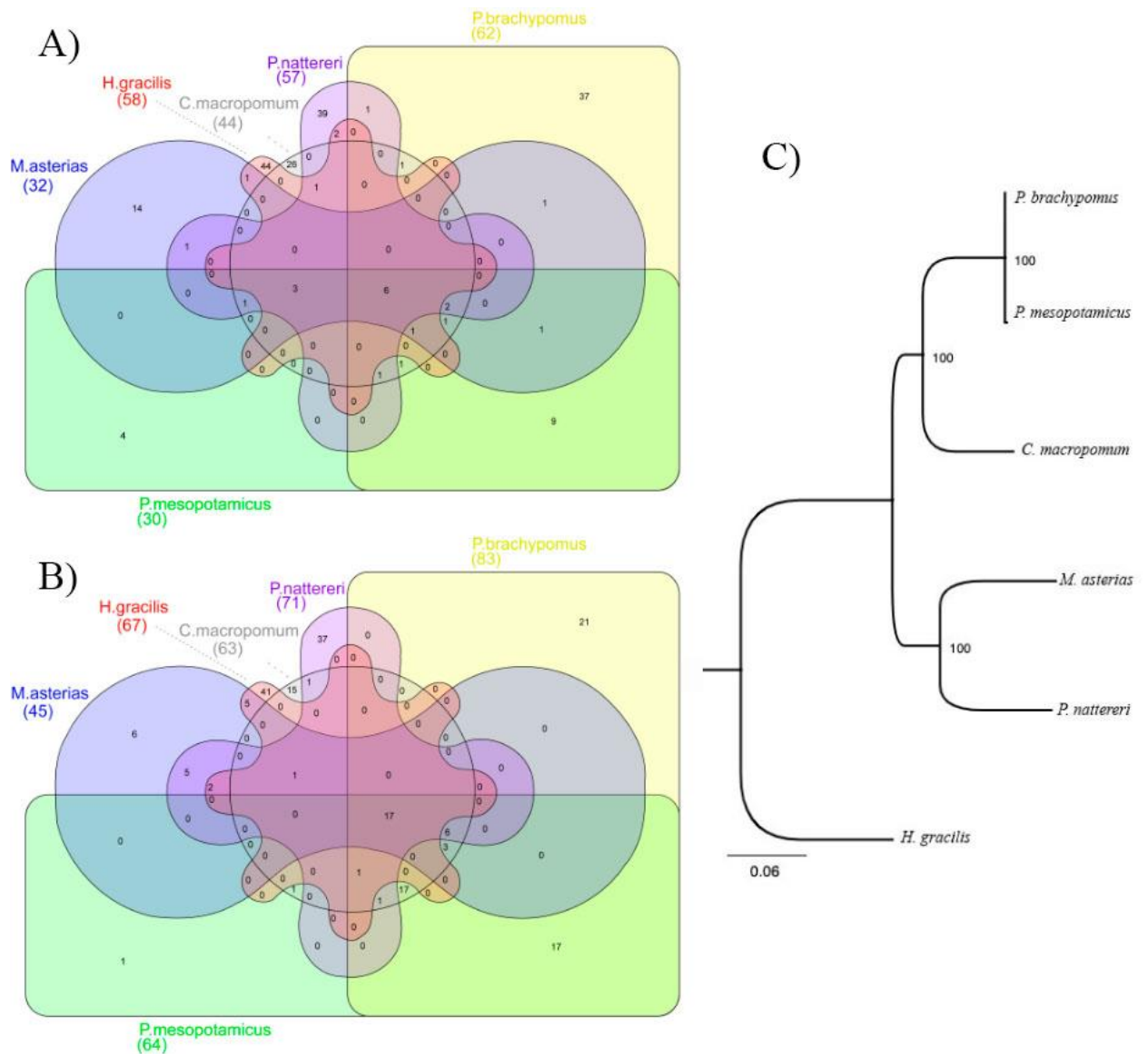


Figura 1: Diagrama de Venn mostrando o número de famílias de satDNAs (>80%) compartilhadas entre as seis espécies após cada nível de análise. (A) Satélites compartilhados após a caracterização clássica. (B) Satélites compartilhados após a clusterização forçada. (C) Filogenia da família Serrasalminidae.

Os resultados obtidos através do mapeamento de *reads* e visualização gráfica dos perfis de abundância (RepeatProfiler) apresentaram várias sequências com padrões de DNAs satélites, com 97 sequências para *P. mesopotamicus* e *P. brachypomus*, 80 para *C. macropomum*, 78 para *H. gracilis*, 77 para *P. nattereri* e 58 para *M. asterias*. De forma geral, os gráficos de cobertura apresentaram padrões similares, podendo ou não haver variação quantitativa, como por exemplo o PmeSat07 (Figura 2a), presente em todas espécies, apresentou uma cobertura similar nas seis espécies, com um pico próximo a

6000 cópias, enquanto o MasSat16 (Figura 2b), também identificado em todas espécies, apresentou padrões de cobertura variando de acordo com a filogenia do grupo, sendo que dentro do clado Colossomatinae houve variação média de 700 cópias, enquanto entre os clados Mylenae e Serrasalminae 1500 cópias em média, ao ponto que *H. gracilis*, representante da família Hemiodontidae, apresentou uma cobertura de apenas 40 cópias. Os gráficos de perfis de variante também seguiram certos padrões filogenéticos. Apesar de algumas sequências terem apresentado perfis semelhantes entre todas as espécies, como o Pmesat07 (Figura 2c), em sua grande maioria, os perfis apresentaram maior semelhança de acordo com a proximidade filogenética entre os grupos, como o MasSat16 (Figura 2d). Outras sequências apresentaram padrões de repetições distintas aos DNAs satélites aqui descritos, como por exemplo o PbrSat08 (Figura 2e) que se apresentou completo apenas entre *P.mesopotamicus* e *P.brachypomus*, e nas demais espécies, apresentou fragmentos de monômeros conservados. Os gráficos obtidos nesta análise também mostraram que alguns satélites com resquícios identificados via RepeatMasker, não apresentaram representação gráfica como o CmaSat37 (Figura 2f).

A busca por mapeamento de *reads* via RepeatMasker dos DNAs satélites aqui descritos, mostraram que dos 198 DNAs satélites inicialmente catalogados, apenas 38 não apresentaram nenhum mapeamento em todas as espécies, das 146 sequências presentes em pelo menos 1 indivíduo representantes do grupo Colossomatinae, apenas 9 não estiveram presentes nas 3 espécies. Dos 46 DNAs satélites inicialmente identificados apenas em *H. gracilis*, 32 tiveram sinais em todas as espécies, e apenas 5 se mostram exclusivos do grupo.

As MSTs dos três DNAs satélites selecionados aleatoriamente mostraram baixo nível de homogeneização intraespecífico, evidenciando um elevado número de haplótipos (monômeros) compartilhados entre as espécies (Figura 3), revelando um cenário inesperado de acordo com as previsões de evolução molecular de satDNAs. Por outro lado, a MST construída com os satélites identificados via RepeatMasker, apresentou menos monômeros (conforme esperado) e uma maior variação entre monômeros, embora um haplótipo central pôde ser observado (Figura 3c).

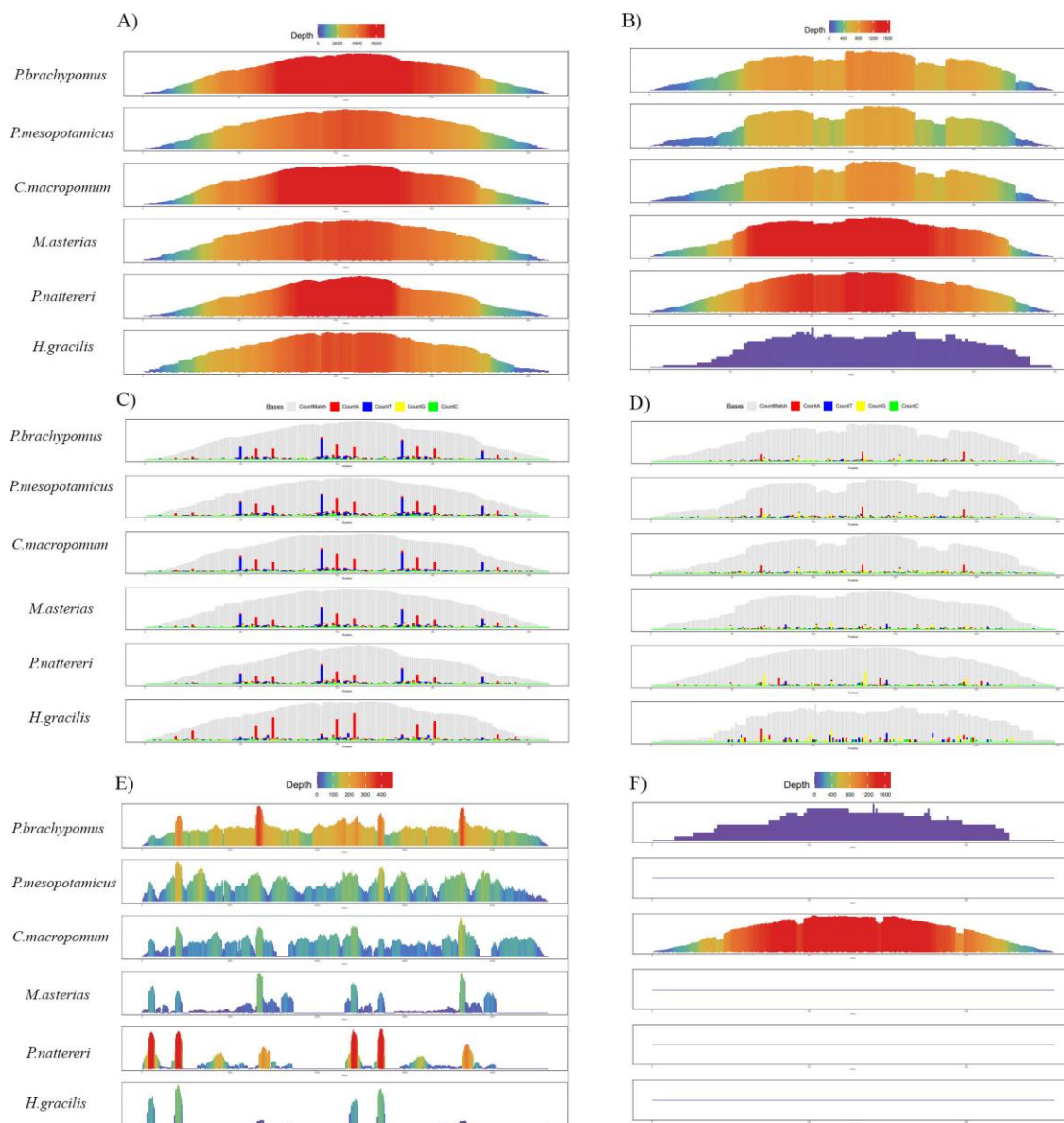


Figura 2: Gráficos RepeatProfiler: (A) PmeSat07 mostrando níveis de cobertura semelhantes em ambas espécies; (B) MasSat16 mostrando variação de níveis de cobertura de acordo com o grupo filogenético; (C) PmeSat07 com perfis de variação semelhantes; (D) MasSat16 com perfis de variação distintos de acordo com o grupo filogenético; (E) PbrSat08 mostrando sequências incompletas, com fragmentos conservados; (F) CmaSat37 identificado em todas espécies via RepeatMasker, com ausência de clusterização via RepeatProfiler

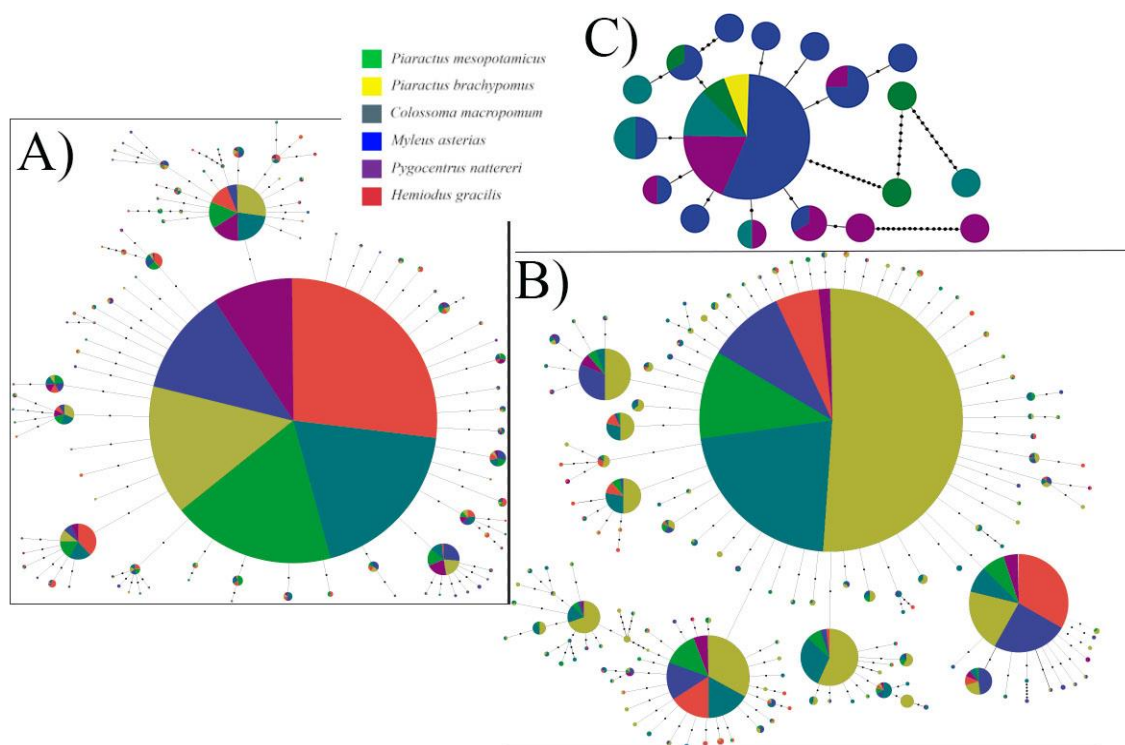
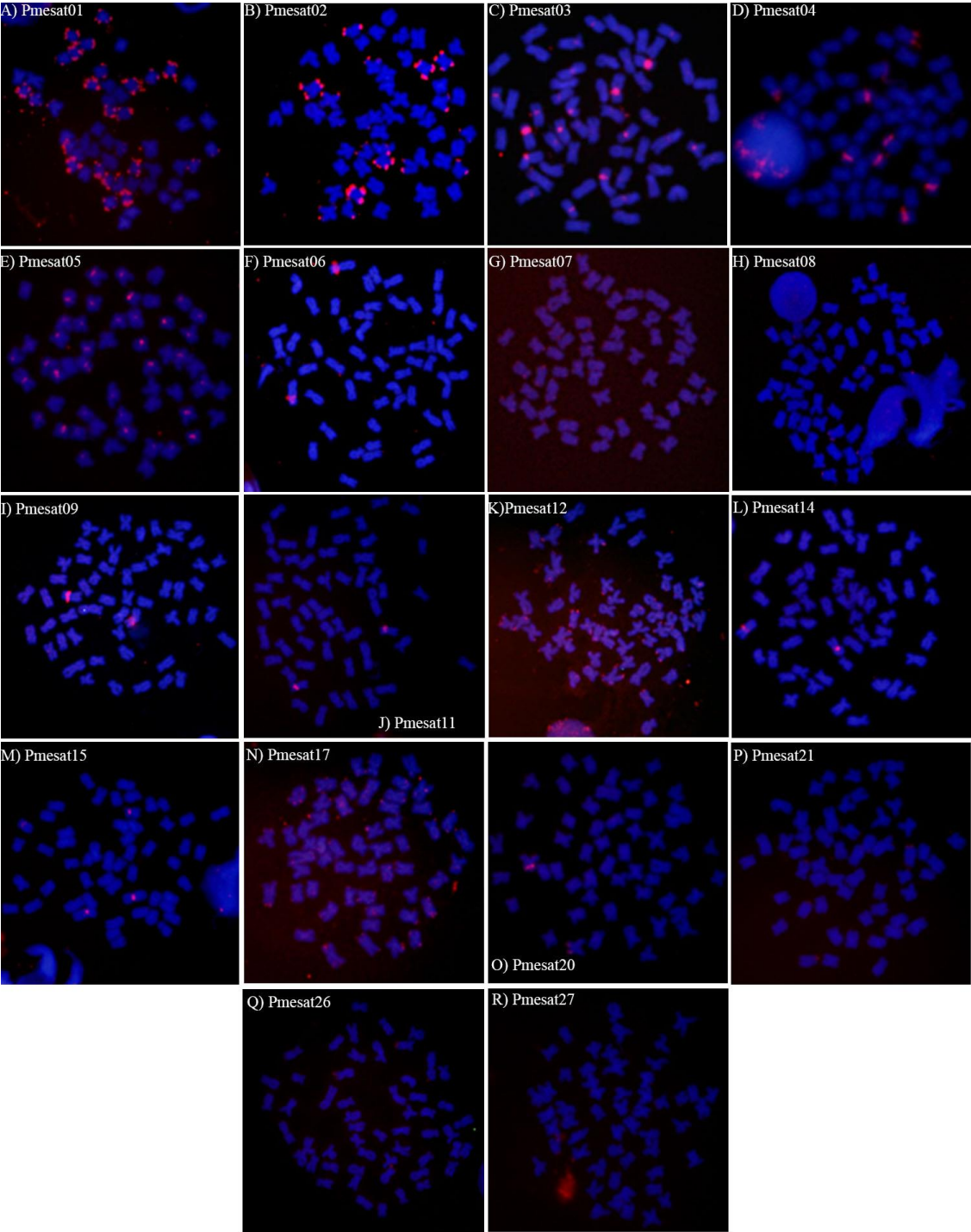


Figura 3: *Minimum spanning trees* de 3 DNAs satélites compartilhados entre as seis espécies analisadas no presente trabalho. (A) PmesSat10; (B) PmesSat19; (C) MasSat16

Dentre os 23 DNAs satélites de *Piaractus mesopotamicus* selecionados para hibridação, 18 apresentaram sinais clusterizados nos cromossomos desta espécie (Figura 4). Os DNAs satélites PmesSat03, 04, 05, 09, 11, 14 e 15, apresentaram acúmulo principalmente nas regiões centroméricas, sendo o PmesSat05 (Figura 4e), presente em quase 100% dos cromossomos. Por outro lado, os DNAs satélites PmesSat01, 2, 6, 7, 8, 12, 20, 21, 26 e 27 apresentaram marcações localizadas principalmente na região telomérica dos cromossomos. O DNA satélite PmesSat17 foi o único com padrões de marcação mista, ocupando tanto a região do centrômero quanto do telômero.



918

919 **Figura 4:** Híbridações *in situ* fluorescência dos DNAs satélites de *Piaractus*
920 *mesopotamicus*

921

922 **4.4 Discussão**

923 Neste estudo, nós caracterizamos os satelitomas de cinco espécies da família
924 Serrasalminidae (*Colossoma macropomum*, *Myleus asterias*, *Piaractus brachypomus*,

Piaractus mesopotamicus e *Pygocentrus nattereri*) e de espécie pertencente à família Hemiodontidae (*Hemiodus gracilis*), sendo compostos por 46, 31, 63, 30, 57 e 59 famílias de DNAs satélites, respectivamente, sendo o satelitoma de *P. mesopotamicus*, o representante com menor número de DNAs satélites entre as espécies de peixes já estudadas até o momento, sendo essas *Psalidodon paranae*, *Characidium gomesi* e *Megaleporinus macrocephalus*, todas pertencentes à ordem dos Characiformes e apresentando 45, 59 e 164 famílias de DNAs satélites caracterizados, respectivamente (Silva et al. 2016; Utsunomia et al. 2019; Serrano-Freitas et al. 2020). Embora os números anteriores de satDNAs em peixes já fossem significativos, nenhuma análise integrativa havia sido realizada, exceto para uma ou duas sequências sabidamente compartilhadas entre um grupo de espécies (Utsunomia et al. 2017; Rodrigues et al. 2019; dos Santos et al. 2021). Nesse sentido, os resultados apresentados aqui não só aumentam significativamente o conhecimento quantitativo de satDNAs em peixes, mas também estabelece um *framework* de análise comparativa que deve ser útil em estudos posteriores.

Os DNAs satélites são sequências dinâmicas e altamente suscetíveis a mutações que se fixam e se espalham rapidamente nas populações, o que leva essas sequências a se tornarem espécie- ou grupo-específicas ao longo do tempo (Garrido-Ramos 2017). No entanto, em alguns casos raros, essas sequências podem se manter conservadas em um grupo de espécies durante um longo período de tempo (Robles et al. 2004; Plohl et al. 2010; Halbach et al. 2020; dos Santos et al. 2021), o que poderia estar relacionado à possível função destas sequências ou a outras causas, como taxas diferenciais de homogeneização (Plohl et al. 2010; Ferreira et al. 2020; Halbach et al. 2020). Notavelmente, a análise comparativa detalhada realizada neste trabalho, mostra um interessante padrão de conservação de repetições *in tandem* entre espécies que divergiram há aproximadamente 60 milhões de anos (Kolmann et al. 2021).

Uma das principais barreiras para os estudos comparativos dos catálogos de DNAs satélites é a metodologia de análise utilizada. Embora a caracterização *de novo* de satDNAs a partir de *short-reads* tenha sido amplamente utilizada em espécies não-modelo (Macas et al. 2010; Novák et al. 2013; Ruiz-Ruano et al. 2016), a determinação da presença/ausência de um determinado DNA satélite em um genoma (processo fundamental para estudos evolutivos) era problemática, uma vez que a não-detecção de um determinado satDNA em uma espécie não implica necessariamente na ausência desta sequência na espécie referida (falso negativo). Particularmente, em Serrasalmídeos, se apenas considerássemos os satDNAs dos catálogos originais, teríamos 6 satDNAs

compartilhados, enquanto as análises subsequentes revelaram que este número é muito maior. Tal padrão está de acordo com a hipótese *library* de evolução a longo prazo de satDNAs, sendo que as mudanças observadas aqui foram majoritariamente quantitativas (Fry e Salser 1977). Além disso, é interessante verificar que a proximidade filogenética entre cada espécie parece relacionada ao número de DNAs satélites compartilhados, embora a análise de correlação não tenha nos mostrado isso.

A busca por sequências específicas utilizando o algoritmo RepeatMasker é comumente utilizada, sendo possível, a identificação de sequências repetitivas em espécies próximas através de um banco de dados (Palacios-Gimenez et al. 2020b). Aqui, as análises de RepeatMasker realizadas utilizando o catálogo completo de sequências (n= 198 satDNAs), mostrou que mesmo não clusterizados, a grande maioria das sequências identificadas estão presentes nos genomas das espécies de alguma forma. No entanto, as análises de RepeatProfiler mostraram que muitas dessas sequências não apresentam padrões de repetição em *tandem* (Figura 3f). O cenário complexo também pode ser analisado no caso do satélite MasSat16, por exemplo, encontrado inicialmente apenas em *M. asterias*, porém com *reads* identificados em todas as espécies (via RepeatMasker). É interessante verificar que a MST obtida para este satDNA (Figura 2.c) apresentou padrões de homogeneização semelhantes aos satélites clusterizados, inclusive com a ocorrência de um monômero conservado e compartilhado. Além disso, esta mesma sequência apresentou padrão de organização de um DNA satélite via RepeatProfiler (Figura 3b, d). Nesse sentido, a não clusterização de todas as sequências identificadas apenas via RepeatMasker podem acontecer por: i) Baixa abundância dessas sequências organizadas *in tandem* no genoma; ii) Disposição isolada no genoma em forma de elementos transponíveis, visto que muitos DNAs satélites são originados a partir dessas sequências, ou mesmo de regiões gênicas (Valeri et al. 2018; Vondrak et al. 2020). Nesse sentido, ressaltamos que a integração de múltiplas ferramentas para a análise comparativa de satDNAs deve ser aplicada para minimizar os impactos das desvantagens de cada tipo de análise isoladamente.

Já é sabido que os DNAs satélites pericentroméricos dos eucariotos podem desempenhar funções ativas nos indivíduos (Jagannathan et al. 2018; Traynor et al. 2019). Assim, o alto número de marcações pericentroméricas do PmeSat05 em *P. mesopotamicus*, além da alta abundância deste mesmo satDNA em *P. brachypomus* (PbrSat07) poderia indicar que este DNA satélite pode desempenhar algum papel importante na organização dos centrômeros dentro do clado dos Colossomatinae, visto

que esta sequência se mostrou exclusiva do grupo, embora análises mais significativas devam ser realizadas sobre este tópico.

Sequências repetitivas compartilhadas entre populações/espécies estão submetidas a mudanças principalmente quantitativas, e como reflexo, com o passar de milhares de anos, acabam passando também por mudanças qualitativas (Fry e Salser 1977), essas mudanças acabam fazendo com que indivíduos próximos apresentem padrões de sequências repetitivas distintas, com monômeros homogeneizados de forma independente em cada espécie (Utsunomia et al. 2017; Rodrigues et al. 2019). Contudo, a análise de MSTs contruídas para dois DNAs satélites aleatórios clusterizados em todas as espécies estudadas (Pmesat10 e 19), e um identificados em todas espécies via RepeatMasker e RepeatProfiler (MasSat16), apresentou um padrão diferente do observado em estudos anteriores para esse tipo de sequência. Aqui, observamos sequências bem conservadas com uma grande quantidade de monômeros compartilhados entre todas as 6 espécies, para os três DNAs satélites. Estes dados, juntamente com a comparação dos monômeros clusterizados por TAREAN e RepeatExplorer2, onde a grande maioria dos DNAs satélites que apresentaram resultados foram classificados pelo menos ao nível de variante (>80%), e a alta similaridade encontrada nos gráficos de RepeatProfiler corroboram com as hipóteses aqui apresentadas de que essas sequências em Serrasalmideos possam estar ligada a uma evolução molecular lenta ou uma possível restrição seletiva.

Com esses resultados, podemos observar um alto nível de conservação de DNAs satélites entre os indivíduos da família Serrasalmidae, seguindo um exemplo interessante da “hipótese *library*” de evolução de satDNAs a longo prazo. Também identificamos “falhas” nas metodologias clássicas de caracterização de satelitos onde o número de satDNAs descritos, não representará a real totalidade dessas sequências dentro do genoma-alvo, podendo gerar resultados contraditórios a depender do tipo de análise comparativa a ser realizada, como por exemplo uma comparação filogenética entre espécies mais distantes ou presença de um DNA satélite específico em determinada espécie. Criando assim, uma metodologia que visa diminuir essa perda de informação.

5. Referências

- Arcila, D., Ortí, G., Vari, R., Armbruster, J. W., Stiassny, M. L., Ko, K. D., Sabaj, M. H., Lundberg, J., Revell, L. J. & Betancur-R, R. (2017). Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology & Evolution*, 1(2), 1-10.
- Baroni, S., Lopes, C. E., & de Almeida-Toledo, L. F. (2009). Cytogenetic characterization of *Metynnis maculatus* (Teleostei; Characiformes): the description in Serrasalminae of a small B chromosome bearing inactive NOR-like sequences. *Caryologia*, 62(2), 95-101.
- Botchan, M. R. (1974). Bovine satellite I DNA consists of repetitive units 1,400 base pairs in length. *Nature*, 251(5473), 288-292.
- Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371(6494), 215-220.
- Cechova, M., Harris, R. S., Tomaszewicz, M., Arbeithuber, B., Chiaromonte, F., & Makova, K. D. (2019). High satellite repeat turnover in great apes studied with short- and long-read technologies. *Molecular biology and evolution*, 36(11), 2415-2431.
- Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic acids research*, 45(4), e18-e18.
- Dos Santos, R. Z., Calegari, R. M., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Melo, S., Oliveira, C., Foresti, F., Uliano-Silva, M., Porto-Foresti, F. & Utsunomia, R. (2021). A long-term conserved satellite DNA that remains unexpanded in several genomes of Characiformes fish is actively transcribed. *Genome Biology and Evolution*, 13(2), evab002.
- Dover, G. A. (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in genetics*, 2, 159-165.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792-1797.
- Eschmeyer, W.N., & Fong, J.D (2020). Species by Family/Subfamily. Eletronic version in <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>
- Escudeiro, A., Adegas, F., Robinson, T. J., Heslop-Harrison, J. S., & Chaves, R. (2019). Conservation, divergence, and functions of centromeric satellite DNA families in the

Bovidae. *Genome biology and evolution*, 11(4), 1152-1165.

Favarato, R. M., Ribeiro, L. B., Ota, R. P., Nakayama, C. M., & Feldberg, E. (2019). Cytogenetic characterization of two *Metynnis* species (Characiformes, Serrasalminidae) reveals B chromosomes restricted to the females. *Cytogenetic and genome research*, 158(1), 38-45.

Ferreira, D., Escudeiro, A., Adegas, F., Anjo, S. I., Manadas, B., & Chaves, R. (2020). FA-SAT ncRNA interacts with PKM2 protein: Depletion of this complex induces a switch from cell proliferation to apoptosis. *Cellular and Molecular Life Sciences*, 77(7), 1371-1386.

Ferreira, M., Garcia, C., Matoso, D. A., de Jesus, I. S., & Feldberg, E. (2016). A new multiple sex chromosome system X 1 X 1 X 2 X 2/X 1 Y 1 X 2 Y 2 in Siluriformes: cytogenetic characterization of *Bunocephalus coracoideus* (Aspredinidae). *Genetica*, 144(5), 591-599.

Flynn, J. M., Lower, S. E., Barbash, D. A., & Clark, A. G. (2018). Rates and patterns of mutation in tandem repetitive DNA in six independent lineages of *Chlamydomonas reinhardtii*. *Genome biology and evolution*, 10(7), 1673-1686.

Foresti, F., & Toledo, L. A. (1981). Polymorphic nature of nucleolus organizer regions in fishes. *Cytogenetic and Genome Research*, 31(3), 137-144.

Fry, K., & Salser, W. (1977). Nucleotide sequences of HS- α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell*, 12(4), 1069-1084.

Garrido-Ramos, M. A. (2015). Satellite DNA in plants: more than just rubbish. *Cytogenetic and Genome Research*, 146(2), 153-170.

Garrido-Ramos, M. A. (2017). Satellite DNA: an evolving topic. *Genes*, 8(9), 230.

Gaviria, J. I., Nirchio, M., Granado, Á., & Estrada, A. (2005). Karyotype and nucleolar organizer regions of *Pygocentrus cariba* (Serrasalminae) from Caicara Del Orinoco, Venezuela. *Interciencia*, 30(1), 44-47.

Halbach, R., Miesen, P., Joosten, J., Taşköprü, E., Rondeel, I., Pennings, B., Vogels, C. B. F., Merklings, S. H., Koenraadt, C. J., Lambrechts, L. & van Rij, R. P. (2020). A satellite repeat-derived piRNA controls embryonic development of *Aedes*. *Nature*, 580(7802), 274-277.

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169 (2015).

- Horz, W., & Zachau, H. G. (1977). Characterization of distinct segments in mouse satellite DNA by restriction nucleases. *European Journal of Biochemistry*, 73(2), 383-392.
- Hutton, J. R., & Wetmur, J. G. (1973). Effect of chemical modification on the rate of renaturation of deoxyribonucleic acid. Deaminated and glyoxalated deoxyribonucleic acid. *Biochemistry*, 12(3), 558-563.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Pesquisa da Pecuária Municipal, disponível em: <https://sidra.ibge.gov.br/tabela/3940#notas-tabela>. Acessado em 30 de novembro de 2020.
- Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M. & Nishida, M. (2013). MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular biology and evolution*, 30(11), 2531-2540.
- J Jagannathan, M., Cummings, R., & Yamashita, Y. M. (2018). A conserved function for pericentromeric satellite DNA. *Elife*, 7, e34122.
- Katoh, K., & Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, 26(15), 1899-1900.
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome research*, 12(4), 656-664.
- Kit, S., & Hsu, T. C. (1961). Relative stability to thermal denaturation of deoxyribonucleic acid (DNA) preparations containing bromodeoxyuridine. *Biochemical and biophysical research communications*, 5(2), 120-124.
- Kopecka, H., Macaya, G., Cortadas, J., Thiéry, J. P., & Bernadi, G. (1978). Restriction enzyme analysis of satellite DNA components from the bovine genome. *European journal of biochemistry*, 84(1), 189-195.
- Kolmann, M. A., Hughes, L. C., Hernandez, L. P., Arcila, D., Betancur-R, R., Sabaj, M. H., López-Fernández, H. & Ortí, G. (2020). Phylogenomics of piranhas and pacus (Serrasalminidae) uncovers how dietary convergence and parallelism obfuscate traditional morphological taxonomy. *Systematic Biology*.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357.
- Lee, H. R., Neumann, P., Macas, J., & Jiang, J. (2006). Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. *Molecular biology and evolution*, 23(12), 2505-2520.

- Liao, D. (1999). Concerted evolution: molecular mechanism and biological implications. American journal of human genetics, 64(1), 24.
- López-Flores, I., & Garrido-Ramos, M. A. (2012). The repetitive DNA content of eukaryotic genomes. In Repetitive DNA (Vol. 7, pp. 1-28). Karger Publishers.
- Lorite, P., Muñoz-López, M., Carrillo, J. A., Sanllorente, O., Vela, J., Mora, P., Tinaut, A., Torres, M. I., & Palomeque, T. (2017). Concerted evolution, a slow process for ant satellite DNA: study of the satellite DNA in the *Aphaenogaster* genus (Hymenoptera, Formicidae). Organisms Diversity & Evolution, 17(3), 595-606.
- Macas, J., Neumann, P., Novák, P., & Jiang, J. (2010). Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. Bioinformatics, 26(17), 2101-2108.
- Macas, J., Kejnovský, E., Neumann, P., Novák, P., Koblížková, A., & Vyskot, B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. PLoS One, 6(11), e27335.
- Miga, K. H. (2020). Centromere studies in the era of ‘telomere-to-telomere’genomics. Experimental cell research, 112127.
- Nakayama, C. M., Feldberg, E., & Bertollo, L. A. C. (2012). Karyotype differentiation and cytotaxonomic considerations in species of Serrasalminidae (Characiformes) from the Amazon basin. Neotropical Ichthyology, 10(1), 53-58.
- Negm, S., Greenberg, A., Larracuente, A. M., & Sproul, J. S. (2021). RepeatProfiler: a pipeline for visualization and comparative analysis of repetitive DNA profiles. Molecular ecology resources, 21(3), 969-981.
- Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet., 39, 121-152.
- Novák, P., Neumann, P., & Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC bioinformatics, 11(1), 378.
- Novak, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics, 29(6), 792-793.
- Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic acids research, 45(12), e111-e111.

- Oliveira, C., Almeida-Toledo, L. F., Foresti, F., Britski, H. A., & Toledo-Filho, S. D. (1988). Chromosome formulae of Neotropical freshwater fishes. *Rev. Brasil. Genet*, 11(3), 577-624.
- Ortí, G., Sivasundar, A., Dietz, K., & Jégu, M. (2008). Phylogeny of the Serrasalminae (Characiformes) based on mitochondrial DNA sequences. *Genetics and Molecular Biology*, 31(1), 343-351.
- Otake, K., Ohzeki, J. I., Shono, N., Kugou, K., Okazaki, K., Nagase, T., Yamakawa, H., Kouprina, N., Larionov, V., Kimura, H. & Earnshaw, W. C. (2020). CENP-B creates alternative epigenetic chromatin states permissive for CENP-A or heterochromatin assembly. *Journal of cell science*, 133(15).
- Palacios-Gimenez, O. M., Koelman, J., Palmada-Flores, M., Bradford, T. M., Jones, K. K., Cooper, S. J., Kawakami, T. & Suh, A. (2020). Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats. *BMC biology*, 18(1), 1-21.
- Palacios-Gimenez, O. M., Milani, D., Song, H., Marti, D. A., López-León, M. D., Ruiz-Ruano, F. J., Camacho, J. P. M. & Cabral-de-Mello, D. C. (2020). Eight Million Years of Satellite DNA Evolution in Grasshoppers of the Genus *Schistocerca* Illuminate the Ins and Outs of the Library Hypothesis. *Genome Biology and Evolution*, 12(3), 88-102. A
- Palacios-Gimenez, O. M., Bardella, V. B., Lemos, B., & Cabral-de-Mello, D. C. (2018). Satellite DNAs are conserved and differentially transcribed among *Gryllus* cricket species. *DNA Research*, 25(2), 137-147. B
- Park, S. T., & Kim, J. (2016). Trends in next-generation sequencing and a new era for whole genome sequencing. *International neurology journal*, 20(Suppl 2), S76.
- Petraccioli, A., Odierna, G., Capriglione, T., Barucca, M., Forconi, M., Olmo, E., & Biscotti, M. A. (2015). A novel satellite DNA isolated in *Pecten jacobaeus* shows high sequence similarity among molluscs. *Molecular Genetics and Genomics*, 290(5), 1717-1725.
- Pinkel, D., Straume, T., & Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences*, 83(9), 2934-2938.
- Plohl, M., Petrović, V., Luchetti, A., Ricci, A., Šatović, E., Passamonti, M., & Mantovani, B. (2010). Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity*, 104(6), 543-551.

- Plohl, M., Meštrović, N., & Mravinac, B. (2012). Satellite DNA evolution. In Repetitive DNA (Vol. 7, pp. 126-152). Karger Publishers.
- Plohl, M., Meštrović, N., & Mravinac, B. (2014). Centromere identity from the DNA point of view. *Chromosoma*, 123(4), 313-325.
- Reis, R. E., Albert, J. S., Di Dario, F., Mincarone, M. M., Petry, P., & Rocha, L. A. (2016). Fish biodiversity and conservation in South America. *Journal of fish biology*, 89(1), 12-47.
- Ribeiro, L. B., Matoso, D. A., & Feldberg, E. (2014). Chromosome mapping of repetitive sequences in four Serrasalminae species (Characiformes). *Genetics and Molecular Biology*, 37(1), 46-53.
- Robles, F., de la Herrán, R., Ludwig, A., Rejón, C. R., Rejón, M. R., & Garrido-Ramos, M. A. (2004). Evolution of ancient satellite DNAs in sturgeon genomes. *Gene*, 338(1), 133-142.
- Rodrigues, P. H. D. M., Dos Santos, R. Z., Silva, D. M. Z. D. A., Goes, C. A. G., Oliveira, C., Foresti, F., Porto-Foresti, F. & Utsunomia, R. (2019). Chromosomal and genomic dynamics of satellite DNAs in Characidae (Characiformes, Teleostei) Species. *Zebrafish*, 16(4), 408-414.
- Ruiz-Ruano, F. J., Castillo-Martínez, J., Cabrero, J., Gómez, R., Camacho, J. P. M., & López-León, M. D. (2018). High-throughput analysis of satellite DNA in the grasshopper *Pyrgomorpha conica* reveals abundance of homologous and heterologous higher-order repeats. *Chromosoma*, 127(3), 323-340.
- Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific reports*, 6, 28333.
- Schmieder, R., & Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS one*, 6(3), e17288.
- Serrano-Freitas, É. A., Silva, D. M., Ruiz-Ruano, F. J., Utsunomia, R., Araya-Jaime, C., Oliveira, C., Camacho, J. P. M. & Foresti, F. (2020). Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. *Molecular Genetics and Genomics*, 295(1), 195-207.
- Silva, D. M. Z. A., Daniel, S. N., Camacho, J. P. M., Utsunomia, R., Ruiz-Ruano, F. J., Penitente, M., Pansonato-Alves, J. C., Hashimoto, D. T., Oliveira, C., & Porto-Foresti, F. (2016). Origin of B chromosomes in the genus *Astyanax* (Characiformes, Characidae) and the limits of chromosome painting. *Molecular Genetics and*

Genomics, 291(3), 1407–1418.

Smit, A. F. A., Hubley, R., & Green, P. (2017). 1996–2010. RepeatMasker Open-3.0.

Smith, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science, 191(4227), 528–535.

Spence, J. M., Critcher, R., Ebersole, T. A., Valdivia, M. M., Earnshaw, W. C., Fukagawa, T., & Farr, C. J. (2002). Co-localization of centromere activity, proteins and topoisomerase II within a subdomain of the major human X α -satellite array. EMBO Journal, 21(19), 5269–5280.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic biology, 56(4), 564–577.

Traynor, S., Møllegaard, N. E., Jørgensen, M. G., Brückmann, N. H., Pedersen, C. B., Terp, M. G., Johansen, S., Dejardin, J., Ditzel, H. J., & Gjerstorff, M. F. (2019). Remodeling and destabilization of chromosome 1 pericentromeric heterochromatin by SSX proteins. Nucleic Acids Research, 47(13), 6668–6684.

Utsunomia, R., Ruiz-Ruano, F. J., Silva, D. M., Serrano, É. A., Rosa, I. F., Scudeler, P. E., Hashimoto, D. T., Oliveira, C., Camacho, J. P. M. & Foresti, F. (2017). A glimpse into the satellite DNA library in characidae fish (Teleostei, Characiformes). Frontiers in genetics, 8, 103.

Utsunomia, R., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Goes, C. A. G., Melo, S., Ramos, L. P., Oliveira, C., Porto-Foresti, F., Foresti, F., & Hashimoto, D. T. (2019). Satellitome landscape analysis of Megaleporinus macrocephalus (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. Scientific Reports, 9(1), 1–10.

Vicari, M. R., Nogaroto, V., Noleto, R. B., Cestari, M. M., Cioffi, M. B., Almeida, M. C., Moreira-Filho, O., Bertollo, L. A. C. & Artoni, R. F. (2010). Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. Journal of Fish Biology, 76(5), 1094–1116.

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

Vozdova, M., Kubickova, S., Cernohorska, H., Fröhlich, J., Martínková, N., & Rubes, J. (2020). Sequence analysis and fish mapping of four satellite dna families among

cervidae. Genes, 11(5).

Wei, J., & Peng, Y. (2018, December). Investigation in spatial variations and temporal trends of nine operational satellite-derived aerosol optical depth products. In AGU Fall Meeting Abstracts (Vol. 2018, pp. A51L-2332).

6. Material suplementar

Tabela s1: Lista dos *primers* desenhados e utilizados para amplificação dos DNAs satélites analisados no presente trabalho

Sats	Forward	Reverse
PmeSat01	GGGAGCGTCTTTTTGGTCAT	AGGAGCATTGGAATCCGCTT
PmeSat02	GCAAAGGCGTTCGAAGTGAA	AGGTGGACCTTCATGTAATTATTT
PmeSat03	TTTAGCGAGAGGACACTGG	CCAGGTTCTTGTCTCAAGC
PmeSat04	GCACTACCTCGCCTAAATCA	CCCTCTATTATCCACAGAGG
PmeSat05	GTACAATGAGCTCTCAGTAACA	CACTTGGGTGACTAGACCTG
PmeSat06	CCATTTGCTGCACATCCTGC	CCACTCTGCTATCCCAGTCT
PmeSat07	ACTCACTGGCCACTTTATTAG	GAAGCACAAGGTGGGTGTTT
PmeSat08	TGTCCAAGAAATTCAAATCCATCA	GACATTTTGATGCCTTGGAGAG
PmeSat09	GGTCCAGCCTATTCGGAAC	CGCAAAATTGTCAGGCATGG
PmeSat11	GGCCTGTTTTCTTGCTAATCC	CGTTTGGCAGAGTAGATATTCT
PmeSat12	CCTAATGAGAACTGTAGAACAG	GACTCTAAATGTAGGTATTGTGA
PmeSat13	GTTCTATCTATGTAGCAATTTTACG	GAATATATCTGTAACTTTCTACGA
PmeSat14	GGATGCCCATGAAACTACCC	CAGAGTGATCATCTCTGGCAG
PmeSat15	CTTCCGCGAGGGCTATATTTA	CGACCACAAACCTCATCCCA
PmeSat16	TTAGTAAGGCGTGGAACGA	ACGCATTAGGATCTCGTTCC
PmeSat17	CACTGAGATTAATAACTGATCATC	GCTAAAGCCTGAGTAGACTG
PmeSat18	ATACTGGATTAAAATCACTCCACAA	GTGTGTGTAGTTTTTACAGTCTGA
PmeSat20	CCTCGCCTGTTATTTTCATCTG	CGTTTCGATTGTGGTGAAAACA
PmeSat21	CGCCCTCTAAAGGCGTCATT	GAGCGACGTTTGTATTTGATG
PmeSat24	GGTTGAAAATGAAAATCGGGCT	TTCAACGTCTGTCCAACGTC
PmeSat26	GTAAAGTTGTTATGGTCTCCC	TTAGTCTCTGAGAGGGACTTT
PmeSat27	GGCGATATGGCTTCTACTGT	CCTACGCTTCCTGTAGTGTA
PmeSat28	GACCACCATCCAGTGATCCA	GGCAGCATTTAACCTGGTAG

Tabela S1: Principais características dos 30 satélites identificados para *P. mesopotamicus* através das análises de Satminer e TAREAN.

Sats	RUL	Divergência	Abundância	A+T
PmeSat01	508	8.10	0.00938024564847	0.586614173228
PmeSat02	143	10.74	0.00837050449506	0.559440559441
PmeSat03	1068	11.85	0.00752829571648	0.588014981273
PmeSat04	118	10.75	0.00321929695752	0.627118644068
PmeSat05	247	18.62	0.00192171681118	0.619433198381
PmeSat06	1853	8.38	0.00173876485004	0.561791689153
PmeSat07	42	15.06	0.00112403828288	0.52380952381
PmeSat08	177	14.96	0.000920100454153	0.655367231638
PmeSat09	696	7.88	0.000761637713173	0.666666666667
PmeSat10	21	16.97	0.000699945265908	0.761904761905
PmeSat11	1242	1.82	0.000649832617711	0.585346215781
PmeSat12	72	6.72	0.000531965654574	0.680555555556
PmeSat13	92	4.75	0.000517202820966	0.739130434783
PmeSat14	956	1.48	0.000516145167085	0.632845188285
PmeSat15	157	5.05	0.00051570800348	0.503184713376
PmeSat16	42	12.12	0.000440791860828	0.452380952381
PmeSat17	65	6.26	0.000382465774774	0.630769230769
PmeSat18	67	9.11	0.000368667924422	0.65671641791
PmeSat19	30	10.15	0.000351078636727	0.566666666667
PmeSat20	142	13.53	0.00033793955365	0.570422535211
PmeSat21	54	10.04	0.000319230160131	0.518518518519
PmeSat22	28	10.87	0.000303496299531	0.5
PmeSat23	38	8.25	0.000262868288428	0.631578947368
PmeSat24	83	9.93	0.000238189697859	0.542168674699
PmeSat25	6	23.98	0.000219544770862	
PmeSat26	81	7.68	0.000212368841098	0.641975308642

PmeSat27	102	11.35	0.000190154080428	0.598039215686
PmeSat28	398	6.54	0.000171192864555	0.572864321608
PmeSat29	30	13.14	0.000165872361886	0.6
PmeSat30	33	5.67	6,82E+06	0.636363636364

1278

1279 **Tabela S3:** Principais características dos 46 satélites identificados para *C. macropomum*
1280 através das análises de Satminer e TAREAN.

Sats	RUL	Divergência	Abundância	A+T
CmaSat01	144	6.62	0.0138973781848	0.5625
Cmasat02	543	13.07	0.00534262106699	0.580110497238
CmaSat03	177	4.27	0.00515832279798	0.649717514124
CmaSat04	141	5.22	0.00464981734198	0.574468085106
CmaSat05	247	26.04	0.00447214603145	0.631578947368
CmaSat06	3168	14.69	0.00399020836011	0.570391414141
CmaSat07	525	1.99	0.00318866832067	0.571428571429
CmaSat08	285	8.90	0.00280954742442	0.585964912281
CmaSat09	42	12.92	0.00166286059191	0.547619047619
CmaSat10	43	19.89	0.0015857992685	0.488372093023
CmaSat11	234	5.46	0.00130603538822	0.57264957265
CmaSat12	1846	5.91	0.00103171863519	0.663596966414
CmaSat13	2220	12.00	0.00100770295495	0.60045045045
CmaSat14	170	0.03	0.000957265170423	0.564705882353
CmaSat15	21	17.19	0.000899790364136	0.761904761905
CmaSat16	955	0.73	0.000761711743253	0.632460732984
CmaSat17	1946	7.25	0.000604870328876	0.535457348407
CmaSat18	30	9.28	0.000584418722099	0.566666666667
CmaSat19	38	6.71	0.000552801883318	0.657894736842
CmaSat20	72	3.10	0.000547511227581	0.652777777778
CmaSat21	28	10.39	0.000514748729523	0.535714285714
CmaSat22	68	5.11	0.000499621950424	0.632352941176
CmaSat23	898	12.74	0.00046263412187	0.599109131403
CmaSat24	327	3.34	0.00046092372631	0.525993883792
CmaSat25	6	20.53	0.000460113191856	0.5

CmaSat26	50	4.40	0.000432846381179	0.64
CmaSat27	394	4.28	0.000414915104909	0.555837563452
CmaSat28	183	4.23	0.000355566836959	0.535519125683
CmaSat29	34	17.93	0.000348643721667	0.705882352941
CmaSat30	919	14.39	0.000344462155235	0.67247007617
CmaSat31	54	8.78	0.000341978394701	0.537037037037
CmaSat32	237	16.73	0.000327204725656	0.594936708861
CmaSat33	66	6.47	0.000321991406416	0.651515151515
CmaSat34	101	6.93	0.000318887155379	0.60396039604
CmaSat35	30	11.94	0.000298551853102	0.533333333333
CmaSat36	1250	2.49	0.000248489959924	0.5992
CmaSat37	51	2.66	0.000217900077888	0.627450980392
CmaSat38	30	11.07	0.000195997662399	0.633333333333
CmaSat39	192	7.47	0.000150033404571	0.619791666667
CmaSat40	932	2.00	0.000142098919735	0.510729613734
CmaSat41	263	9.78	0.000140756022415	0.577946768061
CmaSat42	83	10.92	0.000134218390595	0.602409638554
CmaSat43	62	7.20	0.000121626330188	0.612903225806
CmaSat44	324	9.01	0.000118418364313	0.682098765432
CmaSat45	289	4.82	0.000102014034236	0.622837370242
CmaSat46	30	6.48	9,73E+06	0.433333333333

Tabela S2: Principais características dos 63 satélites identificados para *P. brachypomus* através das análises de Satminer e TAREAN.

Sats	RUL	Divergência	Abundância	A+T
PbrSat01	515	2.68	0,012875504	59,00%
PbrSat02	141	6.87	0,009422533	56,70%
PbrSat03	1068	8.66	0,007667539	58,60%
PbrSat04	118	12.11	0,005386788	58,50%
PbrSat05	2846	18.21	0,00386478	55,10%
PbrSat06	4663	7.28	0,003396685	58,40%

PbrSat07	247	14.07	0,003183446	60,70%
PbrSat08	2316	15.35	0,002315818	50,60%
PbrSat09	1853	7.59	0,002256969	56,20%
PbrSat10	1938	4.25	0,002003409	66,70%
PbrSat11	805	17.26	0,001736843	60,00%
PbrSat12	42	13.01	0,001582115	54,80%
PbrSat13	1144	10.68	0,001460463	59,40%
PbrSat14	145	1.69	0,001429465	50,30%
PbrSat15	1205	14.75	0,001301884	58,90%
PbrSat16	1886	17.43	0,000831961	59,10%
PbrSat17	709	13.08	0,000802942	51,30%
PbrSat18	246	14.88	0,00079504	64,20%
PbrSat19	72	3.42	0,000758758	66,70%
PbrSat20	3441	10.31	0,000711926	51,00%
PbrSat21	1242	0.87	0,000665974	58,50%
PbrSat22	1045	20.21	0,000613969	58,50%
PbrSat23	6	12.45	0,000611479	50,00%
PbrSat24	119	19.81	0,000600364	63,00%
PbrSat25	68	4.52	0,000585416	63,20%
PbrSat26	28	7.94	0,000528041	50,00%
PbrSat27	142	12.45	0,000511717	57,70%
PbrSat28	959	0.42	0,000485928	63,90%
PbrSat29	1362	2.77	0,000485051	52,30%
PbrSat30	1657	9.14	0,000481648	55,30%
PbrSat31	280	4.25	0,000470939	55,70%
PbrSat32	66	5.99	0,000469611	65,20%
PbrSat33	54	13.80	0,000452272	53,70%
PbrSat34	1229	7.50	0,000451901	55,50%
PbrSat35	246	16.89	0,00044252	60,60%
PbrSat36	2088	12.64	0,000439853	54,60%
PbrSat37	2110	0.85	0,000408955	55,20%
PbrSat38	102	6.18	0,000375784	59,80%
PbrSat39	1036	20.41	0,000349877	44,40%
PbrSat40	398	4.73	0,00029613	55,80%

PbrSat41	2778	6.40	0,000268741	50,20%
PbrSat42	427	15.65	0,000253935	61,10%
PbrSat43	91	4.73	0,00025105	73,60%
PbrSat44	1125	1.72	0,000244063	59,60%
PbrSat45	48	6.35	0,00022996	64,60%
PbrSat46	1465	12.07	0,000228783	51,80%
PbrSat47	97	4.98	0,000224946	59,10%
PbrSat48	929	7.24	0,0002025	68,70%
PbrSat49	1124	2.10	0,000192386	52,50%
PbrSat50	247	15.36	0,000186513	61,50%
PbrSat51	184	13.74	0,000175016	55,40%
PbrSat52	83	9.63	0,000168597	55,40%
PbrSat53	1184	8.50	0,000165996	58,60%
PbrSat54	638	11.03	0,000155304	65,70%
PbrSat55	1793	2.17	0,000148322	52,00%
PbrSat56	33	2.00	0,000132779	60,60%
PbrSat57	83	9.91	0,000125882	59,00%
PbrSat58	43	5.51	0,000122367	65,10%
PbrSat59	83	5.68	0,000108886	53,00%
PbrSat60	42	7.96	9,59E-05	59,50%
PbrSat61	142	9.26	9,31E-05	68,30%
PbrSat62	576	5.55	9,02E-05	64,60%
PbrSat63	1499	3.42	6,56E-05	59,80%

1286

1287 **Tabela S4:** Principais características dos 31 satélites identificados para *M. asteris* através
1288 das análises de Satminer e TAREAN.

Sats	RUL	Divergência	Abundância	A+T
MasSat01	246	9.78	0.0155750399132	0.617886178862
MasSat02	73	4.48	0.00200919623041	0.643835616438
MasSat03	42	13.30	0.00159889180317	0.547619047619
MasSat04	30	11.85	0.00143213262398	0.566666666667
MasSat05	418	3.19	0.00136684182804	0.61004784689
MasSat06	67	6.95	0.0012986534446	0.65671641791
MasSat07	52	9.75	0.00118314836272	0.365384615385

MasSat08	54	8.57	0.00110620835952	0.518518518519
MasSat09	63	14.88	0.00104325072519	0.650793650794
MasSat10	68	11.45	0.000866701624048	0.661764705882
MasSat11	102	7.42	0.000754205807515	0.588235294118
MasSat12	177	9.73	0.000724463279112	0.621468926554
MasSat13	21	17.10	0.000675853230408	0.761904761905
MasSat14	28	13.25	0.000585006316348	0.464285714286
MasSat15	177	12.78	0.000579585413059	0.661016949153
MasSat16	62	11.46	0.000480060828089	0.661290322581
MasSat17	946	1.61	0.000293946669866	0.627906976744
MasSat18	6	24.82	0.00029065217324	0.5
MasSat19	38	12.68	0.000288804961206	0.631578947368
MasSat20	40	7.24	0.000269474128736	0.625
MasSat21	142	7.78	0.00025827012372	0.584507042254
MasSat22	30	11.79	0.000201628324652	0.366666666667
MasSat23	30	12.63	0.000194070450578	0.633333333333
MasSat24	34	21.37	0.000173960891467	0.705882352941
MasSat25	34	8.11	0.000164721812977	0.558823529412
MasSat26	36	12.16	0.000148410809984	0.527777777778
MasSat27	54	11.27	0.000146118395703	0.444444444444
MasSat28	83	6.89	9,54E+05	0.578313253012
MasSat29	150	3.19	9,45E+06	0.54
MasSat30	97	6.46	9,37E+06	0.59793814433
MasSat31	130	2.15	2,15E+06	0.592307692308

Tabela S5: Principais características dos 57 satélites identificados para *P. nattereri* através das análises de Satminer e TAREAN

Sats	RUL	Divergência	Abundância	A+T
PnaSat01	247	23.60	0.00639413828292	0.603238866397
PnaSat02	1960	9.58	0.00505243437666	0.635714285714
PnaSat03	1779	2.53	0.00171284044825	0.544125913435
PnaSat04	1810	16.12	0.00141639064446	0.525966850829
PnaSat05	42	13.35	0.00118144756624	0.547619047619

PnaSat06	177	13.06	0.00104869608468	0.677966101695
PnaSat07	2469	5.33	0.000882880195128	0.609153503443
PnaSat08	65	3.93	0.000763801506872	0.630769230769
PnaSat09	21	18.18	0.000753238255879	0.761904761905
PnaSat10	66	4.12	0.00070673942789	0.651515151515
PnaSat11	72	2.94	0.000703410753169	0.625
PnaSat12	921	1.58	0.000541162980583	0.640608034745
PnaSat13	83	3.76	0.000515326735239	0.650602409639
PnaSat14	1712	4.09	0.000429971990068	0.579439252336
PnaSat15	30	10.18	0.000411457973958	0.533333333333
PnaSat16	1953	0.97	0.000403214319672	0.521249359959
PnaSat17	94	13.11	0.000389570387708	0.606382978723
PnaSat18	6	24.19	0.000380841435309	0.5
PnaSat19	101	4.22	0.000379701946788	0.574257425743
PnaSat20	917	1.85	0.000364375505508	0.529989094875
PnaSat21	236	10.10	0.000328235761244	0.529661016949
PnaSat22	791	6.80	0.00032102470162	0.561314791403
PnaSat23	50	15.04	0.00028659953483	0.52
PnaSat24	284	1.11	0.000273725238639	0.489436619718
PnaSat25	30	13.41	0.00027083055298	0.533333333333
PnaSat26	62	13.18	0.00026646607678	0.661290322581
PnaSat27	66	10.70	0.000265659028156	0.515151515152
PnaSat28	44	8.74	0.000247890062233	0.568181818182
PnaSat29	1299	1.43	0.000233278740859	0.554272517321
PnaSat30	27	10.23	0.000217623066497	0.481481481481
PnaSat31	53	16.29	0.000216525266581	0.679245283019
PnaSat32	698	9.82	0.000210522107547	0.448424068768
PnaSat33	948	1.67	0.000204548878793	0.607594936709
PnaSat34	115	4.29	0.000199615727272	0.460869565217
PnaSat35	52	7.11	0.000190974427836	0.634615384615
PnaSat36	51	9.34	0.00017692643709	0.56862745098
PnaSat37	1020	1.62	0.00017365762293	0.527450980392
PnaSat38	36	5.60	0.000172913641682	0.583333333333
PnaSat39	444	6.30	0.000165164905956	0.497747747748

PnaSat40	1121	4.53	0.000164087415873	0.540588760036
PnaSat41	185	7.37	0.000157616061743	0.578378378378
PnaSat42	30	12.95	0.000150108906134	0.633333333333
PnaSat43	530	1.29	0.000145064585001	0.547169811321
PnaSat44	980	1.78	0.000143575553567	0.562244897959
PnaSat45	785	1.77	0.000140102572139	0.51847133758
PnaSat46	276	8.36	0.000126485363639	0.554347826087
PnaSat47	566	4.27	0.000123202653277	0.657243816254
PnaSat48	654	12.37	0.00011333955705	0.703363914373
PnaSat49	42	13.66	0.000106980441471	0.619047619048
PnaSat50	31	6.56	0.000103256259477	0.741935483871
PnaSat51	267	7.90	0.000102738893207	0.554307116105
PnaSat52	46	5.03	9,65E+06	0.565217391304
PnaSat53	392	6.83	9,33E+05	0.563775510204
PnaSat54	18	7.77	9,02E+06	0.5
PnaSat55	58	7.26	7,94E+06	0.706896551724
PnaSat56	48	4.71	5,40E+06	0.604166666667
PnaSat57	63	7.42	4,94E+05	0.492063492063

Tabela S6: Principais características dos 59 satélites identificados para *H.gracilis* através das análises de Satminer e TAREAN

Sats	RUL	Divergência	Abundância	A+T
HgrSat01	569	3.46	0.00920622483749	0.579964850615
HgrSat02	236	10.00	0.00413189732566	0.627118644068
HgrSat03	59	7.45	0.00355544251478	0.559322033898
HgrSat04	613	7.50	0.00301409247217	0.535073409462
HgrSat05	461	5.79	0.00200522325208	0.570498915401
HgrSat06	42	9.40	0.00192286117467	0.571428571429
HgrSat07	60	2.85	0.00187306019085	0.683333333333
HgrSat08	236	11.06	0.00142073361841	0.64406779661
HgrSat09	6	16.75	0.0013135623539	0.5
HgrSat10	21	15.49	0.00112916522872	0.761904761905
HgrSat11	65	5.65	0.000870170792452	0.661538461538
HgrSat12	66	1.66	0.000857851073788	0.651515151515

HgrSat13	91	5.68	0.000744231889718	0.593406593407
HgrSat14	94	3.24	0.000664598668575	0.627659574468
HgrSat15	30	11.71	0.000634120867705	0.566666666667
HgrSat16	54	9.57	0.000556308598716	0.592592592593
HgrSat17	15	8.77	0.000535599125922	0.6
HgrSat18	246	6.51	0.000474340032157	0.59756097561
HgrSat19	75	7.46	0.000416355051514	0.426666666667
HgrSat20	126	7.71	0.000413196676875	0.404761904762
HgrSat21	511	11.99	0.000410917914718	0.391389432485
HgrSat22	59	14.34	0.000384987350298	0.508474576271
HgrSat23	59	17.44	0.000384848464117	0.525423728814
HgrSat24	888	1.48	0.000374066781311	0.613738738739
HgrSat25	51	12.10	0.000333023342806	0.647058823529
HgrSat26	32	14.89	0.00033193797302	0.53125
HgrSat27	288	2.85	0.000278603107481	0.520833333333
HgrSat28	59	17.39	0.000278543952255	0.474576271186
HgrSat29	30	9.49	0.000270527133242	0.633333333333
HgrSat30	1215	1.60	0.000269037964743	0.550617283951
HgrSat31	41	12.11	0.000259429098579	0.658536585366
HgrSat32	911	4.86	0.000240514858275	0.43358946213
HgrSat33	958	1.81	0.000227971378542	0.442588726514
HgrSat34	29	14.47	0.000220751869087	0.758620689655
HgrSat35	20	15.13	0.000217228275231	0.6
HgrSat36	1080	8.35	0.000200523353994	0.480555555556
HgrSat37	49	8.28	0.000197046055532	0.448979591837
HgrSat38	288	14.75	0.000195171092086	0.638888888889
HgrSat39	49	10.21	0.000188087896846	0.448979591837
HgrSat40	22	6.78	0.000174315017212	0.590909090909
HgrSat41	638	9.02	0.000169327974521	0.405956112853
HgrSat42	216	5.25	0.000165295131335	0.481481481481
HgrSat43	1050	1.46	0.00016230136254	0.501904761905
HgrSat44	72	7.72	0.000157792705585	0.680555555556
HgrSat45	717	2.26	0.00015581229152	0.41980474198
HgrSat46	33	11.67	0.000143207084594	0.575757575758

HgrSat47	882	3.21	0.000141759067557	0.502267573696
HgrSat48	402	3.46	0.000136777168799	0.504975124378
HgrSat49	100	11.99	0.000136450529077	0.46
HgrSat50	41	7.15	0.000128845224674	0.682926829268
HgrSat51	266	2.70	0.000118791408336	0.612781954887
HgrSat52	516	1.17	0.000117168497589	0.501937984496
HgrSat53	454	3.08	0.00011308164311	0.422907488987
HgrSat54	38	12.34	9,12E+06	0.578947368421
HgrSat55	364	2.25	9,00E+06	0.543956043956
HgrSat56	513	3.83	8,66E+05	0.446393762183
HgrSat57	37	2.61	7,68E+05	0.567567567568
HgrSat58	272	6.01	7,59E+06	0.621323529412
HgrSat59	201	6.29	7,50E+06	0.63184079602

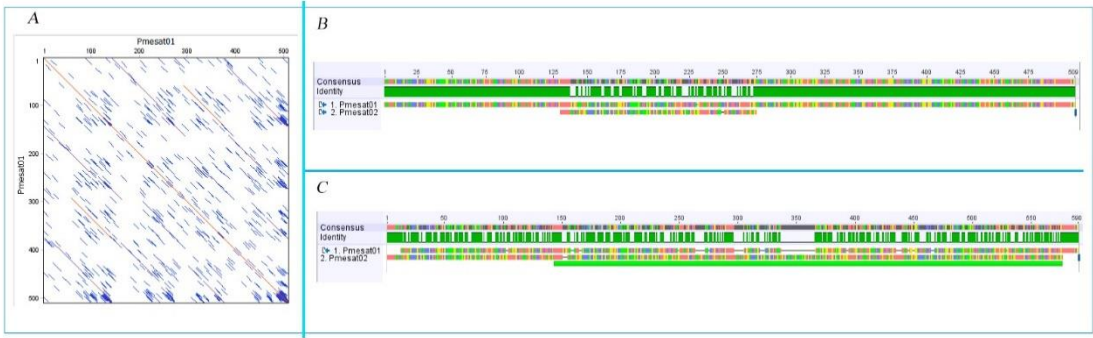
1296

1297 Tabela S7: Satélites compartilhados via caracterização clássica

Pacu	pirapi	tamba	myl	pygo	hemi
12	19	20	2	11	44
18	32	33	6	10	11
27	38	34	11	19	14
17	25	22	10	8	7
7	12	9	3	5	6
21	33	31	8	x	16
29	x	38	23	42	29
22	26	21	14	30	x
19	x	18	4	15	15
10	x	15	13	9	10
1	1	x	x	x	x
14	28	16	17	12	x
28	40	27	x	53	x
3	3	x	x	x	x
x	48	30	x	x	x
8	x	3	15	6	x
2	2	4	x	x	x
x	47	x	30	x	x

30	56	x	25	x	x
x	x	35	x	25	26
6	9	x	x	x	x
x	49	x	x	40	x
11	21	36	x	x	x
x	x	x	1	x	18
15	14	x	x	x	x
5	7	x	x	x	x
20	27	x	x	x	x
4	4	x	x	x	x
13	43	x	x	x	x
24	59	x	x	x	x
x	57	42	x	x	x
x	x	x	x	31	25
x	x	x	16	26	x
x	x	x	x	49	31

Figura s1: Conjunto de imagens evidenciando a existência de uma HOR. (A) *dotplot* mostrando a existência de um conjunto de monômero separado por uma inserção. (B) alinhamento entre PmeSat01 e PmeSat02 mostrando alto nível de homologia. (C) alinhamento de um monômero de PmeSat01 com um trîmero de PmeSat02, corroborando com o observado em (A).



7. Considerações finais

A comparação entre espécies de uma mesma família, a partir de seus satelitomas, mostrou-se eficiente até certo ponto, apesar de relatar de forma correta os níveis filogenéticos entre cada espécie, estas análises podem apresentar uma perda de informação considerável causadas pelas limitações na identificação *de novo* de DNAs satélites em novas espécies. Nesse sentido, a aplicação de uma metodologia que utilize um catálogo geral de sequências repetitivas para buscar estas sequências em outros genomas pode fornecer respostas um pouco mais adequadas, em termos de persistência evolutiva de sequências repetitivas *in tandem*.

Afim de realizar uma caracterização ainda mais completa, as ferramentas bioinformáticas RepeatMasker e RepeatProfiler mostraram-se como ótimos recursos complementares para análises evolutivas e estruturais dessas sequências dentro dos genomas. Essas ferramentas também apresentaram sinais de presença para novos DNAs satélites, porém a utilização desses resultados para definição de presença/ausência, ainda requer mais estudos, devido à falta de visualização concreta da real organização dessas sequências no genoma.

O presente trabalho descreve uma metodologia mais completa para a identificação e comparação de DNAs satélites em um grupo de espécies com grande período de diversificação, tendo apresentado resultados satisfatórios nas análises do grupo Serrasalminidae, onde o resultado obtido em cada uma das etapas está diretamente relacionado a relação filogenética de cada espécie aqui utilizada, indo de acordo com a “hipótese *library*”.