

ANÁLISES CROMOSSÔMICAS E GENÔMICAS APLICADAS
AO ESTUDO EVOLUTIVO E ESTRUTURAL DOS
SATELITOMAS EM ESPÉCIES DO GÊNERO *Gymnotus*
(TELEOSTEI, GYMNOTIFORMES)

SILVANA DE MELO

BOTUCATU - SP

2020



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



UNIVERSIDADE ESTADUAL PAULISTA
"Júlio de Mesquita Filho"
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

ANÁLISES CROMOSSÔMICAS E GENÔMICAS APLICADAS
AO ESTUDO EVOLUTIVO E ESTRUTURAL DOS
SATELITOMAS EM ESPÉCIES DO GÊNERO *Gymnotus*
(TELEOSTEI, GYMNOTIFORMES)

NOME DA CANDIDATA: SILVANA DE MELO

ORIENTADOR: FAUSTO FORESTI

CO-ORIENTADOR: RICARDO UTSUNOMIA

Tese apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

BOTUCATU - SP

2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Melo, Silvana de.

Análises cromossômicas e genômicas aplicadas ao estudo evolutivo e estrutural dos satelitomas em espécies do gênero *Gymnotus* (Teleostei, Gymnotiformes) / Silvana de Melo. - Botucatu, 2020

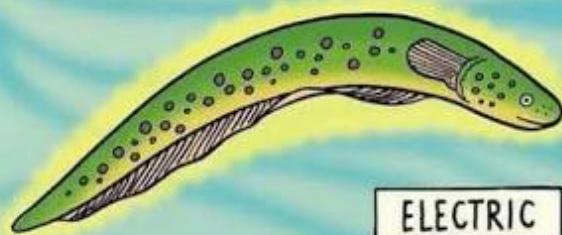
Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Fausto Foresti
Coorientador: Ricardo Utsunomia
Capes: 20203004

1. Citogenética animal. 2. Teleosteos. 3. Peixe - Genética. 4. DNA satélite.

Palavras-chave: Citogenética de peixes; DNA repetitivo; Gymnotidae; Satelitoma.

KNOW YOUR EELS



ELECTRIC



ACOUSTIC

@Wayno®
12-16-16
GOCOMICS.COM/WAYNOVISION

Agradecimentos

“É junto dos ‘bão’ que a gente fica ‘mió’”. (Guimarães Rosa, em “Grande Sertão Veredas” - 1956). Início os agradecimentos com muita alegria pelo o que foi construído ao longo de todos esses anos, felizmente nunca sozinha. Desde a tia Penha, minha primeira professora, deixo aqui os meus agradecimentos a todos os educadores e pensadores que passaram pelo meu caminho e despertaram em mim o interesse pela ciência.

Ao meu orientador, meu pai científico, professor Fausto Foresti, por fazer a utopia de chegar ao final do Doutorado satisfeita com a minha pesquisa se tornar realidade. Não tem palavra que o descreva melhor do que “incentivador”. Muito obrigada, meu professor!

Ao professor Claudio, por toda a ajuda nos momentos pontuais de desespero, que não foram poucos, e todo o aconselhamento ao longo do caminho.

Ao meu co-orientador, professor Ricardo Utsunomia, pela amizade de todos esses anos, oportunidade e suporte científico para a realização desse trabalho.

I also would like to thank professor Nathan Lovejoy for the opportunity to go abroad and study at the University of Toronto, accomplishing a very special chapter in this thesis. This opportunity meant a lot for me, especially for the friends I made. I am so grateful to JP, Frances, Kavi, Lisa, Ahmed, Alex, Katherine and all the undergrads at Lovejoy's Lab.

Aos Desaplaudidos do PDSE na *U of T*, companheiros de perrengues, muitas vezes linguísticos, desafios e conquistas. Especialmente ao Bruno, Áurea, Emanuel e ao combo *cariño* Vanessa e Dario. Aurinha, estou esperando meu cubo mágico.

Aos amigos que fiz durante o PDSE no Canadá. Especialmente à Elaine, Júlia, Kate, Ivan, Leonardo e Helena. Os meus dias foram muito mais lindos e grande parte da saudade que tenho de Toronto está relacionada a vocês.

A minha família LBGP, lugar de tanto apoio e tanto suporte ao longo desses anos. Aos amigos Dani, Ivis, Fabi, Sova, Naná, Lori, Gi, May, Gabi, Bia Boza, Bia Dorini, Angélica, Najila, Marionete, Luz, Bruno Melo, casal Bruno Ota e Rafa Ota, Bruno Campos, Vanessa, Mateus, Cristiano, Cristhian, Paty, Cris, Vivi, Pri, Érica, Li. Especialmente aos integrantes do grupo citogenético *Los Transposons*, Renato, Du e menino Morms. Vocês são meu abrigo, família a gente também pode escolher e vocês são a minha.

À galera do LAGENPE em Bauru, famoso padrão LAGENPE, pelos biscoitos, cafés e análises de quartas esporádicas.

Aos amigos que me apoiaram e caminharam comigo até aqui. Jor, AnaJú, Bia Pereira, Inacio, Danusy, Jú Chiti, Vet, Paula, Carla, Naty Sanches, Marina Oliveira, Marininha.

Aos amigos da pós-graduação, especialmente aos colegas da PG Genética por todos os momentos de entrosamento, ajuda e cuidado.

Aos meus pais e minhas irmãs maravilhosas. Dona Neia, obrigada pelo carinho, consultas podológicas e cafés sempre quentinhos. Sr. Chacal, sou muito grata pelo zelo, pelo apoio emocional e financeiro desde o início. Sinceramente, não teria mesmo como dar esse passo sem o suporte de vocês. Obrigada por sempre acreditarem em mim.

Agradecimento Especial

Gostaria de agradecer especialmente ao ensino público e expressar o meu apoio pela manutenção do mesmo, lugar de excelência em ensino, pesquisa, atividades de extensão e um direito de todos. “O fim do direito é a paz [para o Estado], o meio de que se serve para consegui-lo [o direito] é a luta.” (Rudolf von Ihering, em “A Luta pelo Direito” - 1872).

À CAPES (Processos 88882.433922/2019-01 e 88881.188817/2018-01) pelos auxílios concedidos a mim, às agências de fomento CNPq e FAPESP por auxílios de outrora que possibilitaram a obtenção e manutenção da estrutura laboratorial essencial para o desenvolvimento dessa tese. Estendo, também, agradecimentos especiais ao Programa de Pós-Graduação em Ciências Biológicas (Genética), bem como aos coordenadores e responsáveis pelo funcionamento da Secretaria do PPG.

The analyses that supported the second chapter were funded by Natural Sciences and Engineering Research Council of Canada (discovery grant to Nathan Richard Lovejoy).

Resumo

Um dos tipos mais abundantes de DNA repetitivo, o DNA satélite, apresenta importantes aspectos estruturais e evolutivos, como o acúmulo diferencial entre espécies próximas e taxas de evolução molecular variadas. O satelitoma, identificado como o conjunto total de DNAs satélites (satDNA) de determinado organismo, permite uma investigação mais aprofundada em abundância e riqueza desses elementos repetitivos entre espécies, mas ainda é de conhecimento restrito entre espécies de peixes. Mapeamentos prévios de DNA repetitivo no gênero *Gymnotus* evidenciaram uma ampla diversificação na distribuição dessas sequências no genoma do grupo, todavia é restrita a porção de DNA repetitivo reportado pela limitação de metodologias passadas. Dessa forma, a associação recente de plataformas de Sequenciamento de Próxima Geração (NGS) com análises de bioinformática proporcionam um acesso mais refinado ao genoma, permitindo a anotação de um maior número de elementos repetitivos nos componentes desse grupo. Assim, o objetivo do presente estudo foi investigar a evolução do DNA satélite e a distribuição física de famílias de satDNAs em espécies de peixe elétrico do gênero *Gymnotus*, a partir da construção do satelitoma dos grupos. Os genomas das espécies dos principais clados de *Gymnotus* foram acessados e a metodologia de filtragem pela plataforma RepeatExplorer e satMiner permitiu a identificação do satelitoma em cada um dos grupos abordados, caracterizando massivamente as famílias de satDNAs dentro de complexo de espécies diferentes em *Gymnotus*, destacando os processos de diferenciação genética, a evolução e a dinâmica destas sequências entre representantes do grupo.

Abstract

One of the most relevant types of repetitive DNA, the satellite DNA, has important structural and evolutionary aspects, such as differential accumulation between close species and diverse molecular evolution rates. The satellitome, identified as the total set of satellite DNAs (satDNA) of a given organism, allows further investigation into the abundance and richness of these elements between species, but is still of limited knowledge among fish species. Previous studies of repetitive DNA mapping in the Gymnotidae fish family showed a wide diversification in the distribution of these sequences in the genome of the group. Since past methodologies have been limited in reporting this portion of repetitive DNA, there is still too much to know regarding these elements. Notably, the recent association of Next Generation Sequencing (NGS) platforms with bioinformatics analysis provides a more refined access to the genome, allowing annotation of a greater number of repetitive elements. Thus, the aim of the present study was to investigate the evolution of satellite DNA and the physical distribution of satDNA families in *Gymnotus*, through the construction of the satellitome of the groups. The genomes of the species from the main *Gymnotus* clades were accessed and the RepeatExplorer filtering methodology allowed the identification of the satellitome in each of the groups analyzed, massively characterizing the families of satDNAs within each species complex in *Gymnotus*, highlighting the processes of genetic differentiation, evolution and dynamics of these sequences among representatives of this group.

Sumário

1	INTRODUÇÃO	7
1.1	Estrutura e organização do genoma repetitivo	7
1.2	<i>Next-generation sequencing</i> (NGS) como ferramenta para análises globais de sequências repetitivas	8
1.3	Análise <i>high-throughput</i> e a busca por DNA satélite	9
1.4	Citogenética de peixes e o estudo do DNA satélite	10
1.5	Sistemática e citogenética na ordem Gymnotiformes e no gênero <i>Gymnotus</i>	11
1.6	Justificativa	13
2	OBJETIVOS	14
3	MATERIAIS E MÉTODOS	15
3.1	Amostragem das espécies de <i>Gymnotus</i>	15
3.2	Obtenção de cromossomos metafásicos mitóticos	15
3.3	Coloração convencional com Giemsa	16
3.4	Hibridação in situ fluorescente – FISH	17
3.5	Extração de DNA genômico	18
3.6	Sequenciamento de Próxima Geração (NGS)	18
3.7	Análises bioinformáticas em servidor remoto	19
3.8	Prospecção de satélites com RepeatExplorer – satMiner	19
3.9	Desenho de primers e preparação das sondas	20
4	RESULTADOS E DISCUSSÃO	22
4.1	Capítulo 1 - Satelitoma comparativo e estrutural nas espécies <i>Gymnotus cuia</i> e <i>Gymnotus sylvius</i> (Gymnotiformes, Gymnotidae)	23

4.2	Capítulo 2 - Análise genômica da evolução de DNAs satélites em complexos de espécies do gênero <i>Gymnotus</i> (Teleostei, Gymnotiformes)	40
5	CONSIDERAÇÕES FINAIS	53
6	REFERÊNCIAS	55
7	MATERIAL SUPLEMENTAR	63
7.1	Material Suplementar – Capítulo 1	63
7.2	Material Suplementar – Capítulo 2	68
8	ANEXOS	73
8.1	Anexo I – Certificado do Protocolo de ética CEUA 974/2017 (IBB/UNESP)	73
8.2	Anexo II - RepeatExplorer utilizado x RepeatExplorer <i>TAREAN</i>	74
9	APÊNDICE	77
9.1	Produção Acadêmica e Científica no Doutorado (2016-2020)	77

1 INTRODUÇÃO

1.1 Estrutura e organização do genoma repetitivo

O genoma dos eucariotos se caracteriza por apresentar sequências de nucleotídeos com arranjos variados, formando dois grandes grupos; o primeiro composto por regiões gênicas, e o segundo representado pelas sequências de DNA repetitivo. As sequências de regiões gênicas geralmente apresentam cópia única e constituem uma porção considerável dos genes funcionais, que geram diversos transcritos com funções variadas. Já as sequências de DNA repetitivo são constituídas por segmentos de diversos tamanhos e naturezas, que se repetem inúmeras vezes ao longo do genoma (Charlesworth et al. 1994).

Particularmente, a porção genômica repetitiva apresenta algumas características comuns, podendo ser ressaltado o seu elevado caráter polimórfico nos genomas e ainda o possível acúmulo em regiões que apresentam baixas taxas de recombinação (Charlesworth et al. 1994; Steinemann & Steinemann 2005). De forma geral, o DNA repetitivo compreende sequências dispersas no genoma, como os elementos transponíveis, ou organizadas *in tandem*, como as sequências de famílias multigênicas e as sequências satélites (Charlesworth et al. 1994; Nei & Rooney 2005).

O DNA satélite (satDNA) constitui uma fração não codificante do genoma dos organismos e consiste em sequências repetidas *in tandem* que podem chegar a formar longos *arrays* de dezenas a milhares de nucleotídeos (Garrido-Ramos 2017). Este tipo de sequência está localizado preferencialmente em regiões pericentroméricas e subteloméricas dos cromossomos, embora também possa ocorrer nas regiões intersticiais (López-Flores & Garrido-Ramos 2012; Plohl et al. 2012). A teoria do “Paradoxo do valor C”, que associa diretamente a complexidade dos organismos à quantidade de DNA genômico, é completamente refutada pela presença de DNAs repetitivos como os DNAs satélites, visto que o genoma dos eucariotos apresenta um variável acúmulo desse tipo de sequência (Thomas Jr. 1971; Garrido-Ramos 2017).

As sequências de DNA satélite são representadas por diferentes famílias que variam nos genomas com relação à localização, constituição, tamanho da unidade de repetição e abundância (Garrido-Ramos 2015, 2017). Recentemente, Ruiz-Ruano et al. (2016a) propuseram um caminho evolutivo para cada família de DNA satélite existente no genoma das espécies eucarióticas. Neste estudo, os autores propõem que o nascimento de um satDNA se daria pela duplicação de alguma sequência em determinado local do genoma, formando um

pequeno *array* de sequências; este *array* seria, então, disseminado para diversos locais do genoma, podendo passar subsequentemente por processos locais de amplificação.

O termo “DNA satélite” apresenta um cunho histórico, pois a descoberta desse tipo de DNA repetitivo se deu por um pequeno pico no perfil de ultracentrifugação em Cloreto de Césio (CsCl) (Kit 1961). Atualmente, a técnica de ultracentrifugação não é mais utilizada para procurar sequências de satDNA, tendo sido substituída por ferramentas mais eficazes e resolutivas, tais como a cinética de renaturação (Britten et al. 1974) e digestão com enzimas de restrição seguida de eletroforese, originando um padrão de escada (Singer 1982). Mais recentemente, foi desenvolvida a metodologia de análise bioinformática de sequências de DNA geradas por Sequenciamento de Próxima Geração (NGS) (Novák et al. 2013). Nesse sentido, a incorporação de ferramentas de bioinformática permitiu uma alta eficiência na detecção de sequências de DNA repetitivo, incluindo DNAs satélites.

1.2 *Next-generation sequencing* (NGS) como ferramenta para análises globais de sequências repetitivas

As tecnologias de NGS, recentemente desenvolvidas para estudos genômicos, tornaram o processo de caracterização de sequências nucleotídicas mais eficiente, com sequenciamentos que geram uma quantidade massiva de dados, tornando possível até mesmo a comparação de genomas inteiros de organismos. Tais avanços metodológicos têm sido acompanhados por novas propostas de análises bioinformáticas com *pipelines* amigáveis, possibilitando o estudo de determinadas porções do genoma dos organismos.

Análises prévias de regiões repetitivas de DNA a partir de dados providos da aplicação de técnicas de NGS tem se constituído um desafio nos últimos anos, principalmente porque a maioria das ferramentas de bioinformática geralmente requer genomas previamente montados para a identificação da repetição, ou baseiam-se em buscas de similaridade em bancos de dados de elementos repetitivos já conhecidos (Novak et al. 2010). Contudo, tais particularidades dificultam o estudo e a caracterização destas regiões em organismos não-modelos. Nessa abordagem, para que as técnicas funcionem com sucesso, é necessário que os organismos em estudo possuam genomas bem montados, o que ainda não é comum em espécies de peixes neotropicals.

Novak et al. (2010) descreveram uma abordagem diferencial para a identificação de repetições utilizando os *reads* providos diretamente do sequenciamento *de novo* do DNA. Esta abordagem consistiu em demonstrar, de forma gráfica (*graph-based*), grupos de *reads* que se

sobrepõem frequentemente formando *clusters*. Tal estratégia foi utilizada com sucesso para a análise da porção repetitiva do genoma em plantas (Novak et al. 2010; Macas et al. 2011) e animais (Pagan et al. 2012; Ruiz-Ruano et al. 2016b; Silva et al. 2017; Utsunomia et al. 2016, 2019). Além disso, destaca-se que o número de *reads* gerados pelo sequenciamento na abordagem citada mostrou-se proporcional à abundância genômica das sequências analisadas (Macas et al. 2007; Novak et al. 2010). Assim, a eficiência deste método para o estudo da composição repetitiva presente em cromossomos poderia apontar tanto as repetições mais abundantes como aquelas exclusivas nas diferentes bibliotecas, tal como demonstrado em plantas do gênero *Rumex* (Steflova et al. 2013). Dessa forma, a aplicação da técnica de NGS aliada à análise de *reads in silico*, poderia identificar e caracterizar de forma mais rápida e massiva as famílias de DNA repetitivo acumuladas em cromossomos específicos ou espécies, em comparação com as informações obtidas por métodos clássicos, realizados com base em cortes no genoma com enzimas restrição.

1.3 Análise *high-throughput* e a busca por DNA satélite

O sequenciamento massivo é tido como uma metodologia de alto rendimento, podendo gerar milhões de *reads* a custo moderado comparado à plataformas anteriores, o que contribuiu positivamente para estudos de DNA satélites reportados recentemente na literatura (Ruiz-Ruano et al. 2016a, 2016b; Palacios-Gimenez et al. 2017; Utsunomia et al. 2016, 2019). Adicionalmente às milhões de sequências geradas após o sequenciamento (*throughput*), novas *pipelines* de análise computacional (*in silico*) estão sendo desenvolvidas para a análise desses *reads* e filtragem otimizada do DNA satélite, sendo possível construir uma biblioteca com todos os satDNAs presentes no genoma dos organismos (Ruiz-Ruano et al. 2016a, 2016b; Silva et al. 2017; Utsunomia et al. 2016, 2019).

Estudos pioneiros permitindo o acesso a toda coleção de satDNA foram realizados no gafanhoto *Locusta migratoria*, no qual 62 famílias de sequências foram anotadas (Ruiz-Ruano et al. 2016a, 2016b). Posteriormente, foi reportado que uma sequência, dentre toda a coleção de DNA satélites, é responsável por 55% do conteúdo genético do cromossomo B dessa espécie (Ruiz-Ruano et al. 2018). Antes de 2016, as metodologias disponíveis de isolamento por enzimas de restrição e eletroforese permitiram o isolamento de 21 famílias de satDNA em insetos da ordem Orthoptera, em sua maioria espécies de gafanhotos; sendo assim, a nova metodologia por filtragem representa, então, um aumento substancial no conhecimento de sequências satélites (Ruiz-Ruano et al. 2016b).

Entre os peixes, os primeiros estudos reportados em literatura foram nas espécies de Characiformes *Astyanax paranae*, *Megaleporinus macrocephalus* e *Characidium gomesi* com a identificação de 45, 164 e 59 famílias de DNA satélite, respectivamente (Silva et al. 2017; Utsunomia et al. 2019; Serrano-Freitas et al. 2020). Deve-se destacar que nos estudos acima citados, diferentes catálogos de DNA satélite foram descritos e caracterizados, mas nenhuma comparação interespecífica com mais de um genoma analisado foi realizada, de modo a investigar as possíveis relações existentes, bem como a evolução e a dinâmica destas sequências em um grupo de espécies relacionadas.

1.4 Citogenética de peixes e o estudo do DNA satélite

A fauna de peixes de água doce na região continental Neotropical é representada por cerca de 5.160 espécies, constituindo aproximadamente um terço de toda a riqueza de espécies de peixes de água doce mundial (Reis et al. 2016). A identificação e descrição de novas espécies de peixes reportadas em literatura é crescente, sendo estimada a ocorrência de mais de 8.000 espécies na região Neotropical (Reis et al. 2016). No entanto, embora apenas uma pequena parcela destas espécies tenha sido geneticamente analisada, uma intensa variabilidade intra- e interespecífica vem sendo descrita nos últimos 40 anos, incluindo a ocorrência de variações cariotípicas numéricas e estruturais, além da caracterização de cromossomos com características particulares, como os cromossomos supranumerários e certos heteromorfismos ligados ao sexo (revisões em Oliveira et al. 2009; Arai 2011). Com o advento e a popularização da técnica de Hibridação *in situ* Fluorescente (FISH), no final de década de 90, os estudos citogenéticos em peixes se voltaram para a caracterização e mapeamento físico de diversos tipos de DNA repetitivo, especialmente de sequências de DNAs ribossômicos 5S e 18S, além de diferentes satDNAs (Mestriner et al. 2000; Artoni et al. 2006; Martins et al. 2006; Vicari et al. 2010; Cioffi et al. 2011; Melo et al. 2017; Silva et al. 2017).

Embora sejam abundantes no genoma dos peixes, poucos estudos foram realizados até o momento relacionados à caracterização e ao aspecto evolutivo de um conjunto de DNAs satélites considerando diferentes espécies de um grupo biológico (De La Herrán et al. 2001; Lanfredi et al. 2001; Robles et al. 2004; Martins et al. 2006; Lima et al. 2017; Palacios-Gimenez et al. 2017; Silva et al. 2017; Utsunomia et al. 2017, 2019). Além disso, grande parte dos estudos desenvolvidos teve como objetivo principal a geração de marcadores citogenéticos para análises mais refinadas sobre a origem e evolução de cromossomos supranumerários e de cromossomos relacionados ao sexo (Koehler et al. 1997; Mestriner et al. 2000; Jesus et al. 2003; Artoni et al.

2006; Parise-Maltempo et al. 2013; Utsunomia et al. 2016; Melo et al. 2017; Silva et al. 2017). Deve ser destacado que a técnica de restrição enzimática do genoma para a caracterização destas sequências foi utilizada em alguns dos estudos citados acima. No entanto, esta metodologia apresenta a eficiência limitada, pois utiliza do princípio do acaso para que as enzimas selecionadas cortem o DNA na região desejada do satDNA a ser estudado. 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 o estudo sistematizado de um mesmo satDNA, ou de uma coleção deles (satelitoma), em diferentes espécies (Vicari et al. 2010; Ruiz-Ruano et al. 2016a, 2016b; Garrido-Ramos 2017; Utsunomia et al. 2017). Nesse sentido, metodologias como NGS associadas a análises de bioinformática, que permitem o isolamento, caracterização e análise de sequências repetitivas distribuídas por todo o genoma são extremamente relevantes, considerando-se especialmente de regiões de DNA satélite, tornando possível a construção da biblioteca de satDNA a partir do isolamento de todo o conjunto de DNA satélites existentes no genoma analisado.

1.5 Sistemática e citogenética na ordem Gymnotiformes e no gênero *Gymnotus*

A ordem Gymnotiformes é representada por cerca de 219 espécies válidas distribuídas em cinco famílias de peixes elétricos de água doce exclusivos da região Neotropical (Albert & Crampton 2003; Tagliacollo et al. 2016). A presença de órgãos elétricos nesta ordem de peixes deu origem à emissão de dois tipos de sinais elétricos que evoluíram monofileticamente, dividindo-se em dois grupos principais: Sinusoidea e Pulseoidea (Alda et al. 2019). Os grupos de peixes que apresentam sinais elétricos do tipo "onda" (Sinusoidea) são representadas pelas famílias Sternopygidae e Apterontidae, enquanto o sinal do tipo "pulso" (Pulseoidea) está presente nas famílias Rhamphichthyidae, Hypopomidae e Gymnotidae (Mago-Leccia 1994; Tagliacollo et al. 2016). Esse grupo de peixes, conhecido como *knifefishes*, apresenta ampla distribuição geográfica nas Américas (com exceção do Chile e Belize), ocorrendo na região cis- e trans- Andina na América do Sul, nas planícies dos pampas argentinos (36° S) até o Rio San Nicolás (região de Chiapas), no sul do México (18° N), sendo encontrados também na Ilha de Trinidad (Mago-Leccia 1994; Albert 2001).

Dentre os Gymnotiformes, a diversidade cariotípica é mais bem caracterizada nos gêneros *Gymnotus* (família Gymnotidae) e *Eigenmannia* (família Sternopygidae), por serem os táxons mais especiosos, além de estarem amplamente distribuídos pela região Neotropical,

especialmente em bacias da região da Amazônia, Orinoco e Guiana (Albert 2001; Albert & Crampton 2003; Lovejoy et al. 2010).

A família Gymnotidae é atualmente constituída por dois gêneros: *Electrophorus* e *Gymnotus*. *Electrophorus* foi, por décadas, tido como monotípico, com apenas uma espécie válida. Entretanto, duas novas espécies foram descritas recentemente, totalizando três espécies (Santana et al. 2019). Por outro lado, o gênero *Gymnotus* constitui um táxon muito especioso, com 43 espécies válidas atualmente, e muitos táxons aguardando descrição formal (Crampton & Hopkins 2005; Lovejoy et al. 2010; Crampton et al. 2013, Craig et al. 2018, 2019). Por conta da problemática em descrever novas espécies, devido a conflitos relacionados aos caracteres morfológicos, a sistemática do gênero *Gymnotus* é bastante complexa, com relações filogenéticas conflitantes entre alguns táxons e muitas espécies alocadas em complexos de espécies, como por exemplo o complexo *Gymnotus carapo*, situação que subestima a biodiversidade dentro deste gênero (Lovejoy et al. 2010). Notavelmente, os estudos citogenéticos, somados às características morfológicas, têm contribuído de forma significativa para acessar a diversidade presente em *Gymnotus*, especialmente no grupo *G. carapo* (Milhomem et al. 2008, 2012a, 2012b; Utsunomia et al. 2018). A família Gymnotidae possui um número diploide basal de 52 cromossomos (Silva et al. 2019), entretanto as espécies componentes do gênero *Gymnotus* apresentam uma ampla plasticidade cariotípica, com número diploide variando de 34 cromossomos em *G. capanema* a até 54 cromossomos em algumas populações de *G. carapo* e em *G. cuiá* (Utsunomia et al. 2018). A alta similaridade morfológica entre as espécies dentro de um mesmo complexo, como *G. carapo*, juntamente com a alta plasticidade cromossômica, reforça a abordagem citogenética como uma ferramenta eficaz para contribuir com diagnósticos taxonômicos neste grupo de peixes e inclusive sugere alguns marcadores de DNA repetitivo, como o DNAr 5S, como um possível marcador biogeográfico para o gênero (Fernandes-Matioli & Almeida-Toledo 2001; Scacchetti et al. 2011; Milhomem et al. 2012a; Silva et al. 2019).

Os dados citogenéticos disponíveis para o gênero *Gymnotus* refletem a existência de uma ampla e interessante diversidade cariotípica dentro do gênero, relacionada a eventos de rearranjos cromossômicos e a uma acumulação diferencial na distribuição do DNA repetitivo (Silva et al. 2019). No entanto, deve-se destacar que algumas porções do genoma dos representantes deste gênero ainda permanecem desconhecidas, como as sequências de DNA satélite. Nesse sentido, a utilização de múltiplas abordagens, como o Sequenciamento de Próxima Geração e o consequente mapeamento citogenético de regiões satélites a partir do sequenciamento, possibilitariam a obtenção de informações de interesse para a melhor

compreensão da organização do genoma e estrutura cariotípica neste grupo de organismos, aumentando o conhecimento sobre os processos envolvidos na diversificação e evolução dessas espécies da família Gymnotidae.

1.6 Justificativa

Entre as técnicas utilizadas nos estudos do genoma das espécies, a utilização da técnica de NGS associada às análises computacionais constitui uma ferramenta eficiente e capaz de evidenciar tanto sequências satélites mais abundantes quanto aquelas menos abundantes no genoma das espécies, o que permite a ampliação do conhecimento reportado em literatura acerca de DNA satélites em espécies de peixes neotropicais. Embora sejam abundantes nos genomas dos peixes, poucos estudos foram realizados até o momento relacionados à caracterização e ao aspecto evolutivo de DNAs satélites considerando diferentes espécies de um grupo biológico.

Os complexos de espécies que integram o gênero *Gymnotus* apresentam relacionamentos filogenéticos divergentes, alterando a topologia do nível de parentesco de acordo com o tipo de dado utilizado e sua metodologia, como, por exemplo, abordagens moleculares vs. análises morfológicas (Albert 2001; Lovejoy et al. 2010; Crampton et al. 2013; Tagliacollo et al. 2016). A incorporação de ferramentas potencialmente resolutivas como NGS e análises computacionais permitiria uma nova perspectiva de acesso ao conteúdo genético no gênero, investigando se grupos mais próximos filogeneticamente compartilham um número maior de sequências satélites do que grupos mais distantes, utilizando filogenias previamente publicadas, analisando se a evolução do DNA satélite corrobora a história filogenética do grupo.

Além disso, estudos com base em uma análise comparativa entre bibliotecas de espécies diferentes possibilitariam mostrar o “tempo de vida” dos DNAs satélites, com famílias de satDNA surgindo e desaparecendo ao longo da história evolutiva do grupo, bem como satDNAs ancestrais presentes em todos os genomas dos grupos em análise (Ruiz-Ruano et al. 2016a). Sendo assim, a montagem do satelitoma incluindo na análise os principais grupos do gênero *Gymnotus* seria um processo válido para reconstruir a evolução dos satDNAs dentro do grupo.

2 OBJETIVOS

O objetivo de ordem geral do presente trabalho consistiu em investigar a história evolutiva das famílias de DNA satélite em espécies pertencentes ao gênero *Gymnotus*, através de um estudo global dos satelitomas (anotação de todos os satDNAs presentes no genoma) obtidos a partir da associação do Sequenciamento de Próxima Geração com ferramentas de bioinformática, com o propósito de:

- 1) descrever e comparar o conteúdo total de DNAs satélites entre espécies dentro de um mesmo complexo de espécies, *Gymnotus carapo*, de modo a avaliar a riqueza de sequências de DNAs satélites em duas espécies de ocorrência simpátrica;
- 2) compreender a dinâmica de dispersão e acumulação do DNA satélite em cromossomos de diferentes espécies de peixes elétricos, atestando se determinado satélite se localiza preferencialmente nas mesmas regiões cromossômicas, discutindo a evolução cariotípica do grupo;
- 3) investigar se a evolução do DNA satélite entre grupos do gênero *Gymnotus* acompanha filogenias prévias descritas para o grupo, verificando se espécies filogeneticamente mais próximas apresentam o satelitoma com maior similaridade; sendo analisadas espécies dos principais clados de *Gymnotus*: clado *G. coatesi*, clado *G. tigre*, clado *G. cylindricus* e clado *G. carapo*;
- 4) investigar o padrão de abundância genômica nos diferentes grupos das sequências de satDNAs, acessando o estágio evolutivo destes segmentos genômicos com relação à origem, disseminação e acúmulo no genoma das espécies.

3 MATERIAIS E MÉTODOS

3.1 Amostragem das espécies de *Gymnotus*

As análises moleculares e citogenéticas nas espécies *G. cuia* e *G. sylvius* foram realizadas com indivíduos coletados na cachoeira Véu de Noiva, bacia do Rio Paranapanema, região de Botucatu, e no Rio Campo Novo, bacia do Rio Tietê, em Bauru, estado de 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), autorização 974-CEUA/2017 (Anexo I).

Os exemplares analisados foram anestesiados e dissecados, sendo retirado fragmentos do rim anterior para estudos citogenéticos e fragmentos de tecido muscular extraídos e fixados em etanol 95% para posterior extração de DNA e análises moleculares. Os exemplares utilizados na amostragem foram fixados em formol 10%, conservados em álcool 70% e depositados na coleção de peixes do Laboratório de Biologia e Genética de Peixes, Departamento de Morfologia, UNESP- Botucatu-SP.

As análises moleculares com espécies dos outros clados de *Gymnotus* foram acessadas a partir de amostras da coleção ictiológica do laboratório do professor Nathan Lovejoy, na *University of Toronto* (Tabela 1), em parceria com o *Royal Ontario Museum*.

Tabela 1. Espécies de *Gymnotus* analisadas, com voucher, localidade e clado de acordo com classificação proposta por Tagliacollo *et al.* (2016).

Espécie	Código	Voucher	Localidade	Coordenadas	Clado
<i>Gymnotus cuia</i>	38772	LBP 11152	Córrego Campo Novo, Bauru, Brasil	S 22°23'07" W 49°00'55"	<i>G. carapo</i>
<i>Gymnotus sylvius</i>	33925	LBP_11160	Cachoeira Véu da Noiva - Botucatu, Brasil	S 22° 59'25" W 48°25'40"	<i>G. carapo</i>
<i>Gymnotus cylindricus</i>	10505	UTSC NL	Caño Palma, Costa Rica	N 10°36'49.5" W 83°32'23.7"	<i>G. cylindricus</i>
<i>Gymnotus javari</i>	8612	MUSM44715	Rio Amazonas, Peru	S 04°55.637' W 073°39.234'	<i>G. coatesi</i>
<i>Gymnotus inaequilabiatus</i>	13030	INHS110838	Rio Negro, Uruguai	S 32°58'45.8" W 58°3'47.6"	<i>G. tigre</i>

3.2 Obtenção de cromossomos metafásicos mitóticos

Para obtenção dos cromossomos metafásicos foi utilizado o protocolo proposto por Foresti *et al.* (1981), com adaptações. Esta metodologia se baseia na inibição da polimerização dos microtúbulos pela ação da colchicina, a hipotonização das células em suspensão por solução

de KCl e a fixação celular pelo fixador de Carnoy, na proporção de metanol 3:1 ácido acético. O procedimento utilizado consistiu em:

- 1) injetar solução aquosa de colchicina (0,0125%) na região intra-abdominal, na proporção de aproximadamente 0,1 mL / 10 g de peso do animal;
- 2) manter o peixe em aquário bem aerado, por um período de 40 minutos;
- 3) eutanasiar o animal e, após total inércia do mesmo, iniciar a dissecação e a retirada da porção anterior do rim. Transferir o material para uma placa de Petri contendo 7 mL de solução hipotônica de KCl (0,075M) a 37 graus;
- 4) dissociar o material renal com o auxílio de pinças de dissecação, complementando esse processo com o auxílio de uma seringa de vidro até a obtenção de uma suspensão celular homogênea;
- 5) transferir a suspensão obtida para um tubo de centrífuga e deixar o tubo no interior de uma estufa a 37 °C por 21 minutos;
- 6) retirar o tubo da estufa e adicionar 10 gotas de fixador de Carnoy gelado à suspensão celular; agitar levemente a mistura com uma pipeta Pasteur e deixar repousar por cinco minutos à temperatura ambiente;
- 7) adicionar cerca de 5 mL de fixador à suspensão e novamente agitar a mistura com cuidado; levar à centrífuga (1000 rpm) por 10 minutos;
- 8) retirar o sobrenadante, ressuspender o precipitado em 10 mL de fixador e centrifugar por 10 minutos a 1000 rpm;
- 9) repetir o passo anterior;
- 10) retirar o sobrenadante e diluir o precipitado em aproximadamente 1 mL de fixador;
- 11) pingar o material em lâminas secas previamente limpas;
- 12) deixar as lâminas secar ao ar.

As lâminas contendo as preparações cromossômicas foram estocadas a -20°C para a execução de técnicas citogenéticas.

3.3 Coloração convencional com Giemsa

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

3.4 Hibridação *in situ* fluorescente - FISH

As sondas de DNA satélite utilizadas foram marcadas com digoxigenina-11-dUTP (Roche) e produzidas a partir de sequências satélites amplificadas no genoma de *Gymnotus cuia* e *G. sylvius*. A aplicação da técnica de FISH no material cromossômico foi realizada com base em procedimentos adotados por (Pinkel et al.1986), com algumas modificações:

Primeiro dia – Tratamento pré-hibridação

- 1) lavar as lâminas em solução PBS 1x, durante 5 min;
- 2) incubar as lâminas por 10 min em solução de pepsina 0,005% (em 10mM HCl) a 37 °C;
- 3) lavar em PBS 1x durante 5 min (agitando), a temperatura ambiente;
- 4) fixar em formaldeído 1% em PBS 1x/50mM MgCl₂ durante 10 min a temperatura ambiente;
- 5) lavar em PBS 1x por 5 min. (agitando);
- 6) desidratar as lâminas em série alcoólicas (70, 85, 100 %) gelado, por 5 min cada;
- 7) desnaturar a solução de hibridação a 100°C por um período de 10 min e passá-la imediatamente ao gelo;
- 8) desnaturar o DNA cromossômico nas lâminas com formamida 70% em 2xSSC a 70 °C por 5 min;
- 9) desidratar o material em série alcoólica (70, 85 e 100%) gelado, durante 5 min cada;
- 10) montar cada lâmina com 30 mL de solução de hibridação, cobrir com lamínula e deixar overnight a 37°C em câmara úmida para hibridação da sonda no material cromossômico;

Segundo dia– Lavagem pós-hibridação

- 11) lavar as lâminas em solução 0,4xSSC/0,3% Triton a 73 °C durante 5 min (agitando);
- 12) lavar as lâminas em solução 2xSSC/0,1% Triton a temperatura ambiente, por 30 segundos;
- 13) incubar as lâminas em tampão 5% NFDm/4xSSC por 15 minutos;
- 14) lavar 2 x 5min com Tween 0,5%/4xSSC a temperatura ambiente (agitando);
- 15) incubar cada lâmina com a solução de detecção [100 µL de anti-digoxigenina-rodamina conjugada (0.5 µL anti-digoxi-rodamina/100 µL NFDm) ou FITC-avidina (1 µL FITC-avidina/99 µL NFDm)] durante 1 h em câmara úmida e escura, a temperatura ambiente;

- 16) lavar 3 x 5min com Tween 0,5%/4xSSC, ambiente (Shaker);
- 17) desidratar em álcool 70, 85 e 100%, 5 min. cada (secar);
- 18) aplicar 15 uL de Vectashield com DAPI na lâmina e cobrir com lamínula limpa. Guardar no escuro.

As lâminas foram fotografadas com o auxílio de um fotomicroscópio epifluorescente Olympus BX61 acoplado a uma câmera digital Olympus D80.

3.5 Extração de DNA genômico

O DNA genômico foi obtido a partir de amostras de fígado ou músculo preservadas em etanol utilizando-se os Kits Wizard Genomic DNA Purification (PROMEGA) ou DNeasy Blood & Tissue (QIAGEN), conforme instruções recomendadas, 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%.

3.6 Sequenciamento de Próxima Geração (NGS)

O sequenciamento massivo de DNA pela metodologia de NGS foi realizado com a utilização da plataforma Illumina MiSeq (*paired-end* 2x250pb em *G. cuia* e *G. sylvius* e 2x150 em *G. cylindricus*, *G. inaequilabiatus* e *G. javari*). Não há registro em literatura da existência de heteromorfismo ligado ao sexo ou cromossomos supranumerários (cromossomos B) nas espécies analisadas, dessa forma, foi selecionado o material genômico de um indivíduo por espécie. Análises prévias estimam que o genoma das espécies de *Gymnotus* tenha um conteúdo de DNA de cerca de 1 Giga-base (Gb) (Hinegardner & Rosen 1972). Considerando que a cobertura de 1x do genoma possibilita o acesso a elementos repetitivos e a caracterização dos satelitomas (Ruiz-Ruano et al. 2016a; Utsunomia et al. 2016), cerca de 1 Gb de cada espécie foi sequenciado.

3.7 Análises bioinformáticas em servidor remoto

As análises de bioinformática foram realizadas em servidores remotos da Universidade de Granada e da Universidade Estadual Paulista. O servidor Locusta está localizado no Laboratório de Informática do Departamento de Genética da Universidade de Granada, em Granada, Espanha, sob cuidados do Dr. Juan Pedro Martinez Camacho. O servidor Zungaro está localizado no Instituto de Biociências de Botucatu da Universidade Estadual Paulista, sob

os cuidados do Dr. Claudio de Oliveira. As análises realizadas seguiram um *workflow* padrão para cada uma das bibliotecas sequenciadas utilizando *scripts* disponíveis e descritos em Ruiz-Ruano et al. (2016a).

3.8 Prospecção de DNAs satélites com RepeatExplorer – satMiner

A análise do DNA repetitivo para o estudo das sequências de DNA satélite seguiu um conjunto de ferramentas denominado “satMiner” (Ruiz-Ruano et al. 2016a). Este conjunto de ferramentas consistiu na realização sequencial de diferentes passos: i) preparar os *reads* para aplicação no programa RepeatExplorer (Novak et al. 2013), já com os cortes nos adaptadores e seleção de qualidade; ii) rodar o software RepeatExplorer com a opção padrão; iii) selecionar manualmente, com auxílio do software Geneious Prime 2020, os *clusters* que apresentam a forma de gráficos de DNA satélite (forma em anel ou cilíndrica); iv) montar uma base de dados customizada com os satélites já isolados e utilizar um *script* (*deconseq_run.py*) que eliminará todas as sequências que apresentarem similaridade com os satélites já isolados das bibliotecas originais; v) selecionar novos *reads* para uma nova rodada de RepeatExplorer; e vi) realizar as etapas (i a v) até não ser possível encontrar mais sequências repetidas *in tandem*.

A aplicação desta metodologia possibilita encontrar um grande conjunto de sequências de DNA satélite presentes no genoma das espécies analisadas. Deve ser destacado que o protocolo de satMiner foi aplicado de forma independente em cada uma das bibliotecas sequenciadas, sendo possível realizar a análise comparativa entre as bibliotecas ao final dos procedimentos. Novák et al. (2017) disponibilizaram uma nova plataforma de análise de sequências repetidas *in tandem* (TAREAN), no qual os *reads* que contêm sequências satélites em potencial são isolados de forma automática. Os *clusters* contendo todo tipo de sequência repetida *in tandem* são analisados para a identificação de gráficos de forma circular, selecionando automaticamente *clusters* que contêm pontos fortemente conectados em seus gráficos, diferenciando assim os possíveis satélites de outros elementos repetitivos.

Essa abordagem mais recente chegou a ser considerada no presente estudo, em duas iterações-teste no genoma de *G. sylvius*, entretanto os dados disponibilizados geraram uma quantidade de satDNA inferior à plataforma de Novák et al. (2013), especialmente de satélites com monômeros de menor grandeza, com menos de 165 nucleotídeos, o que poderia futuramente subestimar a real diversidade de satDNAs nas espécies analisadas (ANEXO II).

Após o término do isolamento de todos os satDNAs encontrados, cada satDNA foi comparado contra a coleção inteira de satDNAs utilizando-se o software RepeatMasker v4.0.5

(Smit et al. 2010). Sequências satélites alinhadas (algoritmo Needleman-Wunsch) que apresentaram entre 50% e menos de 80% de similaridade foram consideradas diferentes famílias de satDNA que compartilham uma mesma superfamília; sequências que apresentaram identidade entre 80% e menos de 95%, foram consideradas como variantes de uma mesma família de satDNA; e as que apresentaram a partir de 95% de similaridade, foram consideradas a mesma variante de um mesmo DNA satélite (Ruiz-Ruano et al. 2016a). Após o isolamento do satelitoma de cada uma das espécies amostradas, foi montado um catálogo devidamente nomeado com todas as sequências das espécies analisadas, seguindo Ruiz-Ruano et al. (2016a), com os satélites renomeados de acordo com a ordem decrescente de abundância, seguido por um traço e a indicação do tamanho do monômero. Por exemplo: o satélite ApaSat02-236, descrito em Silva et al. (2017), é o segundo satélite mais abundante de *Astyanax paranae*, com a unidade de repetição composta por 236 nucleotídeos.

3.9 Desenho de *primers* e preparação das sondas

Sequências iniciadoras foram desenhadas para a amplificação de satDNAs selecionados. Os monômeros dos satélites foram alinhados em *De Novo* para a obtenção de sequências consenso e os *primers* de famílias de satélites compartilhados e com considerável abundância foram desenhados manualmente na região mais conservada, sendo obtidos *primers* divergentes (orientação oposta) para satélites com menos de 800 pb e convergentes para satélites com monômeros muito longos (acima de 1000 pb), de modo a otimizar o processo de anelamento durante a amplificação. Os *primers* foram qualitativamente checados com a ferramenta Multiple Primer Analyzer (ThermoFisher Scientific), para evitar a seleção de *oligos* incompatíveis e com alta Temperatura de *Melting* (T_m).

A Reação de Amplificação em Cadeia (PCR) foi realizada a partir do DNA genômico de *G. cuia* e *G. sylvius* em diluições fracionadas (1:10, 1:100, 1:1000, (...), 1:100.000), com os seguintes parâmetros: 8.5 μ L de água autoclavada, 1,25 μ L de tampão da enzima Taq polimerase (10X), 0,75 μ L de $MgCl_2$ (25 mM), 0,25 μ L de dNTP (2 mM), 0,25 μ L de Primer A + Primer B (10 mM), 0,1 μ L de Taq polimerase (5 U/ μ L) e 1 μ L de DNA molde, em um total da reação de 12 μ L. As marcações da sonda seguiram um padrão semelhante, com a substituição do DNA molde pelo produto de PCR, a troca do dNTP regular pelo dATP/dCTP/dGTP (2mM) + dTTP (1mM) e pela incorporação de Digoxigenin-11-dUTP (Roche Applied Science).

As amplificações no termociclador seguiram um programa molde, com redução do tempo de anelamento para 10 segundos para amplificar satélites com monômeros curtos, de acordo com recomendação de Ruiz-Ruano et al. (2016a):

35 ciclos {
95° C -> 5 minutos
95° C -> 20 segundos
55°-60° C -> 40 segundos
68° C -> 20 segundos
68° C -> 7 minutos

A qualidade dos produtos de PCR e das sondas marcadas foram checadas em eletroforese de gel de agarose a 2%.

4 RESULTADOS E DISCUSSÃO

A seção de resultados e discussão obtida através dos estudos genômicos e cromossômicos em espécies do gênero *Gymnotus* se encontra organizada em forma de capítulos, que seguem a formatação de manuscritos científicos.

4.1 Capítulo 1

Satelitoma comparativo e estrutural nas espécies *Gymnotus cuia* e *Gymnotus sylvius* (Gymnotiformes, Gymnotidae)

4.2 Capítulo 2

Análise genômica da evolução de DNAs satélites em complexos de espécies do gênero *Gymnotus* (Teleostei, Gymnotiformes)

Capítulo 1

Satelitoma comparativo e estrutural nas espécies *Gymnotus cuia* e *Gymnotus sylvius* (Gymnotiformes, Gymnotidae)

Resumo

A busca e descoberta de novas sequências de DNA repetitivo, através da associação de técnicas de sequenciamento de genoma com análises computacionais, têm revelado importantes aspectos estruturais e evolutivos desses elementos nos genomas eucarióticos, como por exemplo a sua elevada abundância e dinâmica saltatória. Até o momento, não se tem uma perspectiva comparativa de satelitomas em grupos ou espécies de peixes intimamente relacionados, com mais de um genoma analisado e comparado. Assim, o objetivo do presente trabalho foi investigar a riqueza de DNA satélite em *Gymnotus cuia* e *G. sylvius*, descrevendo os satelitomas dessas duas espécies de ocorrência simpátrica nas bacias dos rios Tietê e Paranapanema. A filtragem de satélites permitiu a detecção de 92 e 69 famílias de DNA satélites em *G. cuia* e *G. sylvius*, respectivamente, com monômeros de extensão e abundância genômica diversificadas nas duas espécies. Os resultados obtidos revelaram que 43 famílias de DNA satélites em *G. cuia* compartilham variados graus de identidade de sequência com satélites detectados em *G. sylvius*. O mapeamento cromossômico de alguns satélites abundantes compartilhados possibilitou inferências acerca da dinâmica estrutural desses elementos entre as duas espécies, que integram o complexo de peixes *Gymnotus carapo*.

Introdução

A identificação de sequências de DNA satélite por muito tempo se ancorou em técnicas que possibilitavam o isolamento aleatório por meio do uso de enzimas de restrição, o que limitava a extensão do número de famílias de satDNAs a serem analisadas (Garrido-Ramos 2017). No entanto, destaca-se que a partir do surgimento e expansão das tecnologias de Sequenciamento de Próxima Geração (NGS), associadas com o desenvolvimento de ferramentas de bioinformática, a investigação da organização de regiões de DNA satélite de diversos organismos não-modelos se tornou possível, por não ser necessário um genoma de

referência para as análises (Novak et al. 2013; Ruiz-Ruano et al. 2016a). Esta recente integralização de técnicas possibilitou a construção do satelitoma em insetos (Ruiz-Ruano et al. 2016a, 2016b), plantas (Fu et al. 2019; Ebrahimzadegan et al. 2019) e peixes (Silva et al. 2017; Utsunomia et al. 2019; Serrano-Freitas et al. 2020). O reconhecimento e anotação de regiões repetitivas em espécies não-modelos é uma abordagem vantajosa para futuros estudos envolvendo a montagem dos genomas dessas espécies, visto que a abundância de elementos repetitivos, juntamente com o alto polimorfismo dessas sequências, tem representado um desafio para a montagem de genomas.

A região das bacias do Rio Paranapanema e Rio Tietê abriga duas espécies simpátricas de peixes Gymnotiformes com ampla distribuição, *Gymnotus cuia* e *Gymnotus sylvius*. Ambas espécies integram o clado *Gymnotus carapo*, táxon mais especioso do gênero *Gymnotus* (Craig et al. 2019). Devido à complexidade em discriminar morfologicamente *G. sylvius* e *G. cuia*, se faz necessário a realização de diagnósticos moleculares para a identificação dessas espécies. A técnica de PCR-RFLP (Faria-Pereira et al. 2019) se mostrou eficaz na diferenciação entre *G. cuia* e *G. sylvius*, entretanto, a utilização de técnicas de citogenética também se apresenta como uma alternativa rápida e viável, visto que essas espécies apresentam um número diploide notavelmente distinto, com *G. cuia* portando 54 cromossomos e *G. sylvius* portando 40 cromossomos.

O mapeamento físico de algumas sequências de DNA repetitivo nessas duas espécies permitiu a elaboração de hipóteses interessantes sobre a dinâmica evolutiva e diferenciação cariotípica no grupo (Scacchetti et al. 2011, 2012; Utsunomia et al. 2014). A distribuição cromossômica do DNAr 18S se mostrou extensamente conservada em diferentes espécies de *Gymnotus carapo* (Milhomem et al. 2013), indicando uma homeologia do locus do DNAr 18S no mesmo par cromossômico. O DNAr 5S, por outro lado, é um potencial marcador biogeográfico de espécies do complexo *Gymnotus carapo* (Silva et al. 2019). Adicionalmente, *G. cuia* e *G. sylvius* apresentam duas classes de DNAr 5S (5S-I e 5S-II), cada uma apresentando um modo particular de distribuição no genoma das espécies, com um restrito grau de dispersão de 5S-I no genoma das duas espécies e um alto grau de dispersão de 5S-II apenas no cariótipo de *G. cuia*, com 14 pares marcados (Scacchetti et al. 2012). A organização de alguns microsatélites em ambas espécies é compartimentalizada (*motifs* de CA, GA e GAG), em formas de blocos, e também pode apresentar associação com famílias multigênicas (*motifs* de GC), sugerindo que o acúmulo de elementos repetitivos não é randômico e que estes elementos genômicos podem ser muito informativos como marcadores para acessar a evolução cariotípica do grupo.

O DNA satélite atua como promotor na evolução e arquitetura genômicas (Louzada et al. 2020). Em contrapartida, é uma das frações menos conhecidas, mesmo apresentando considerável proporção no genoma. Desse modo, a incorporação de metodologias mais refinadas de análise genética seria interessante para evidenciar a riqueza e a distribuição de sequências repetitivas em *Gymnotus*, especialmente de sequências de DNA satélite que ainda não foram descritas para essas espécies.

Metodologia

As análises citogenéticas e genéticas nas espécies *G. cuiia* e *G. sylvius*, foram realizadas com indivíduos coletados na cachoeira Véu de Noiva, bacia do Paranapanema, em Botucatu, e no Rio Campo Novo, bacia do Rio Tietê, em Bauru e seguiram recomendação da Comissão de Ética no Uso de Animais do Instituto de Biociências de Botucatu (Protocolo CEUA 974/2017). Os cromossomos mitóticos foram obtidos a partir de suspensão do rim anterior, fixado com *Carnoy* na proporção metanol 3:1 ácido acético, de acordo com protocolo descrito por Foresti et al. (1981). Como não há registro de cromossomos sexuais descritos para nenhuma dessas espécies, o material genético proveniente de um espécime de *G. cuiia* e de um de *G. sylvius* foi sequenciado, com cobertura aproximada de 1x do genoma (1 Gb) (Hinergadner & Rosen, 1972) realizado através da plataforma Illumina MiSeq (PEx250).

O método de filtragem dos elementos repetitivos foi realizado através da plataforma RepeatExplorer, com o isolamento dos satélites putativos, com gráficos circulares, selecionados manualmente com o uso do software Geneious Prime 2020. Os satélites extraídos foram retirados (“descontaminados”, através do *script* deconseq) da subsequente seleção de *reads* realizada através do pacote de *scripts* satMiner (Ruiz-Ruano et al. 2016a), com a realização de seguidas iterações até que nenhum satélite inédito fosse encontrado, totalizando 12 iterações para *G. cuiia* e oito iterações para *G. sylvius*. Os satélites foram nomeados em ordem decrescente de abundância, seguindo recomendação de Ruiz-Ruano et al. (2016a).

As sondas dos satélites de ocorrência em comum nas duas espécies foram amplificadas a partir de *primers* desenhados em uma região consenso das sequências e foram marcadas com digoxigenina-11-dUTP. A aplicação da técnica de FISH seguiu o protocolo proposto por Pinkel et al. (1986), em altas condições de estringência para evitar sinais inespecíficos e as metáfases foram visualizadas com microscópio epifluorescente Olympus BX61 acoplado à câmera Olympus DP80. As fotos foram registradas com o auxílio do software Olympus cellSens e editadas com o software Adobe Photoshop 2020.

Resultados

Os satDNAs extraídos dos genomas de *G. cuia* e *G. sylvius* apresentaram composição de bases e tamanho de monômeros variados (Tabela 1.1). Em *G. cuia* foi isolado o total de 96 sequências, representando 92 famílias de satDNA (Figura 1.1). Quatro superfamílias foram identificadas no genoma da espécie; a superfamília A', composta pelos satélites GcuSat12 e GcuSat27; a superfamília B', formada pelos satélites GcuSat50 e GcuSat71; a superfamília C', formada por GcuSat65 e GcuSat81; e a superfamília D', composta por GcuSat76 e GcuSat91 (Tabela Suplementar 1.1 e Figura Suplementar 1.1).

A análise do genoma de *G. sylvius* permitiu identificar e isolar 69 famílias de DNA satélite, com o total de 73 sequências (Figura 1.2). Adicionalmente, duas superfamílias de satDNA, A e B, foram detectadas. A superfamília A, formada pelas famílias GsySat18, GsySat20 e GsySat21 e a superfamília B, formada pelos satélites GsySat64 e GsySat69 (Tabela Suplementar 1.2 e Figura Suplementar 1.2).

Seguindo classificação de Ruiz-Ruano et al. (2016a), 43 satDNAs de *G. cuia* apresentaram algum nível de identidade de sequência com os satélites de *G. sylvius* (Tabela 1.2). Quatro sequências de GcuSats apresentaram mais de 50% de similaridade com GsySats (Superfamília), 18 GcuSats apresentaram mais de 80% de similaridade com outros GsySats (Famílias) e 22 GcuSats apresentaram acima de 95% de similaridade com outros GsySats (Variantes). Os satélites com abundância considerável entre as duas espécies foram mapeados a partir de *primers* sintetizados em regiões consenso dos satélites alinhados (Tabela 1.3).

Dentre as sequências compartilhadas, o mapeamento físico identificou padrões interessantes de acúmulo desses elementos repetitivos, com marcações em blocos (clusterizados) ou dispersas ao longo dos cromossomos de *G. cuia* e *G. sylvius* (Tabela 1.4). O satélite GcuSat06-592|GsySat13-706 (71,4% de similaridade) apresentou um padrão mais disperso com alguns pares fracamente clusterizados em *G. cuia* e apenas um bloco conspícuo em *G. sylvius* (Figura 1.3).

Dentre os satélites com mais de 80% de similaridade, a marcação em blocos localizados em regiões pericentroméricas foi predominante (Figuras 1.4, 1.5 e 1.6). Os satélites GcuSat01-118|GsySat03-139 e GcuSat02-611|GsySat01-216 se acumularam preferencialmente na região pericentromérica nas duas espécies, evidenciando uma extensiva acumulação nos cromossomos de ambas as espécies. O satélite GcuSat03-537|GsySat04-1186 apresentou acumulação predominantemente pericentromérica em *G. cuia* e não apresentou nenhuma marcação

relacionada ao par de NOR, já em *G. sylvius* houveram algumas marcações subterminais e o par de NOR marcou na região subterminal e pericentromérica. O satélite GcuSat04-221|GsySat02-602 se organizou extensivamente na região pericentromérica de mais de 10 pares em *G. cuia*, enquanto em *G. sylvius* a sua acumulação se deu na forma de marcações pericentroméricas e intersticiais em quatro pares cromossômicos. O satélite GcuSat47-338|GsySat17-1392 apresentou o mesmo padrão de marcação entre as espécies, tanto em relação ao número de cromossomos marcados quanto à localização física das sequências.

O satélite GcuSat12-181|GsySat07-181 apresentou marcação em blocos por toda a região da NOR nas duas espécies. Adicionalmente, houve acumulação em outro par metacêntrico de *G. cuia*, o que não ocorreu em *G. sylvius*. O satélite GcuSat45-GsySat08 apresentou um padrão mais disperso, com alguns blocos conspícuos no genoma de *G. cuia* e *G. sylvius*.

Alguns satélites com mais de 95% de similaridade apresentaram uma distribuição mais restrita ao longo dos genomas de *G. cuia* e *G. sylvius* (Figura 1.7). O satélite GcuSat20-1205|GsySat19-1204 apresentou um padrão de marcação em poucos cromossomos, mantendo uma distribuição clusterizada no genoma das duas espécies. Já os satélites GcuSat14-1248|GsySat05-1248 e GcuSat45-172|GsySat08-176 apresentaram uma marcação em pontos no genoma de *G. cuia* e moderadamente clusterizada no genoma de *G. sylvius*.

Tabela 1.1. Diversidade geral de DNA satélites encontrados em *G. cuia* e *G. sylvius*, com média de porcentagem A + T, número de famílias e variação dos monômeros.

satDNA – GsySat GcuSat	GcuSat	GsySat
Menor monômero (nt)	34	30
Maior monômero (nt)	2109	1954
Média (nt)	625.9	511.4
A + T (%)	58,6	56,5
Total de satDNAs (com variantes)	96	73

Tabela 1.2. Satélites compartilhados entre *G. cuia* e *G. sylvius*, com porcentagem de identidade de sequência após alinhamento *De Novo* e classificação do nível de identidade proposta por Ruiz-Ruano et al. (2016a). Satélites marcados com asterisco foram mapeados fisicamente.

Sequências compartilhadas	Identidade de sequência (%)	Classificação	Mapeamento físico
GcuSat06-592 GsySat13-706	71,4	Superfamília	*
GcuSat16-1015 GsySat16-810	76	Superfamília	
GcuSat10-1588 GsySat06-1270	72,3	Superfamília	
GcuSat90-493 GsySat47-803	74,6	Superfamília	
GcuSat01-118 GsySat03-139	81,1	Família	*
GcuSat02-611 GsySat01-216	88,5	Família	*
GcuSat03-537 GsySat04-1186	89,2	Família	*
GcuSat04-221 GsySat02-602	92	Família	*
GcuSat08-262 GsySat02-602	93,1	Família	
GcuSat12-181 GsySat07-181	85,6	Família	*
GcuSat15-703 GsySat25-1083	90,8	Família	
GcuSat16-1015 GsySat43-180	85,8	Família	
GcuSat19-64 GsySat66-64	88,1	Família	
GcuSat27-182 GsySat07-181	81,3	Família	
GcuSat33-1037 GsySat34-1036	92,1	Família	
GcuSat47-338 GsySat17-1392	91,5	Família	*
GcuSat50-95 GsySat22-95	91	Família	
GcuSat51-257 GsySat30-257	93,1	Família	
GcuSat53-892 GsySat23-512	90,5	Família	
GcuSat63-726 GsySat45-255	94,2	Família	*
GcuSat61-581 GsySat46-346	94	Família	
GcuSat69-454 GsySat51-1117	85,6	Família	
GcuSat14-1248 GsySat05-1248	97,5	Variante	*
GcuSat20-1205 GsySat19-1204	98,6	Variante	*
GcuSat21-1224 GsySat14-1225	98,8	Variante	
GcuSat24-34 GsySat15-34	100	Variante	
GcuSat30-107 GsySat18-107	95,8	Variante	
GcuSat34-1110 GsySat24-1384	95,9	Variante	
GcuSat35-1992 GsySat36-1954	97,4	Variante	
GcuSat36-42 GsySat35-42	100	Variante	
GcuSat38-147 GsySat32-49	95,9	Variante	
GcuSat40-1511 GsySat28-1568	98,6	Variante	
GcuSat42-107 GsySat18-107	95,5	Variante	
GcuSat45-172 GsySat08-176	98	Variante	*
GcuSat46-1154 GsySat36-1954	98,6	Variante	
GcuSat49-831 GsySat31-809	97,5	Variante	
GcuSat52-768 GsySat39-798	98,2	Variante	
GcuSat55-847 GsySat58-848	97,8	Variante	
GcuSat58-190 GsySat57-166	98,2	Variante	

GcuSat60-756 GsySat49-756	98,2	Variantes
GcuSat62-84 GsySat48-84	100	Variantes
GcuSat70-243 GsySat56-245	96,5	Variantes
GcuSat82-493 GsySat47-803	97,3	Variantes
GcuSat86-1194 GsySat05-1248	97,5	Variantes

Tabela 1.3. Primers de satélites compartilhados entre os genomas de *G. cuia* e *G. sylvius*, com as seqüências de Primer A e Primer B.

Satélites	Seqüência Primer A	Seqüência Primer B
GcuSat06-592 GsySat13-706	TGTAGTGTAGTGGTTATCAC	CAGTAGGTAAGGGGCAGGAA
GcuSat01-118 GsySat03-139	TACGCAAATTGCAAAACACA	TTGAATATTTGGACTTAAAA
GcuSat02-611 GsySat01-216	GTGTGTTATGCACTTTGCTT	GTTTAGATGTTTGCATTATT
GcuSat03-537 GsySat04-1186	CACATGCAGACATTACTACA	CAGTCAGTCAAGTGTCACAA
GcuSat04-221 GsySat02-602	TTATCCAGGAGCTATAGCAG	TCTGTCTGCATATTGCTGAG
GcuSat12-181 GsySat07-181	GCACTAAATTCAGCCACACT	TTTGGGTGAGAGAGTATTCA
GcuSat47-338 GsySat17-1392	TGAAGGACACATTCAGTAG	ATCACAATGGATGCCAGAAG
GcuSat63-726 GsySat45-255	GGTAGTGATTGTTGCCTCTG	TTGCACCTTCTACTGAATCC
GcuSat14-1248 GsySat05-1248	TAACTCTAAATGCCCATGAA	TTTCCTTGCCACTGTCACTT
GcuSat20-1205 GsySat19-1204	TGAGTTAAGCCATGCTAAAC	TTTAAGCAATTCTTCTTGCG
GcuSat45-172 GsySat08-176	CAGTATGTA AAAACCATTTG	GAAGAAAAATTCACTTCACT

Tabela 1.4. Satélites compartilhados e mapeados, com respectivas abundâncias genômicas e padrões de acumulação entre as espécies *G. cuia* e *G. sylvius*.

Satélite	Abundância genômica		Padrão de acumulação	
	<i>G. cuia</i>	<i>G. sylvius</i>	<i>G. cuia</i>	<i>G. sylvius</i>
GcuSat06-592 GsySat13-706	0,045575	0,009514	D am	C ar PC
GcuSat01-118 GsySat03-139	0,134575	0,022923	C ae PC	C ae PC
GcuSat02-611 GsySat01-216	0,133884	0,215465	C ae PC	C ae PC
GcuSat03-537 GsySat04-1186	0,058035	0,021832	C ae PC	C am PC I T
GcuSat04-221 GsySat02-602	0,055741	0,028718	C ae PC	C am PC I
GcuSat12-181 GsySat07-181	0,02227	0,016511	C ar	C ar
GcuSat47-338 GsySat17-1392	0,002766	0,006958	C ar I	C ar I
GcuSat63-726 GsySat45-255	0,001328	0,001721	D P am	C ar I
GcuSat14-1248 GsySat05-1248	0,020118	0,01744	C P am	C D am PC
GcuSat20-1205 GsySat19-1204	0,010466	0,00664	C ar PC	C ar I
GcuSat45-172 GsySat08-176	0,003008	0,015282	D P am	C D am PC

C= clusterizado, D= disperso, P= acumulação em pontos; ae= acumulação extensa, am= acumulação moderada, ar= acumulação restrita; PC= pericentromérico; I= intersticial; T= subterminal

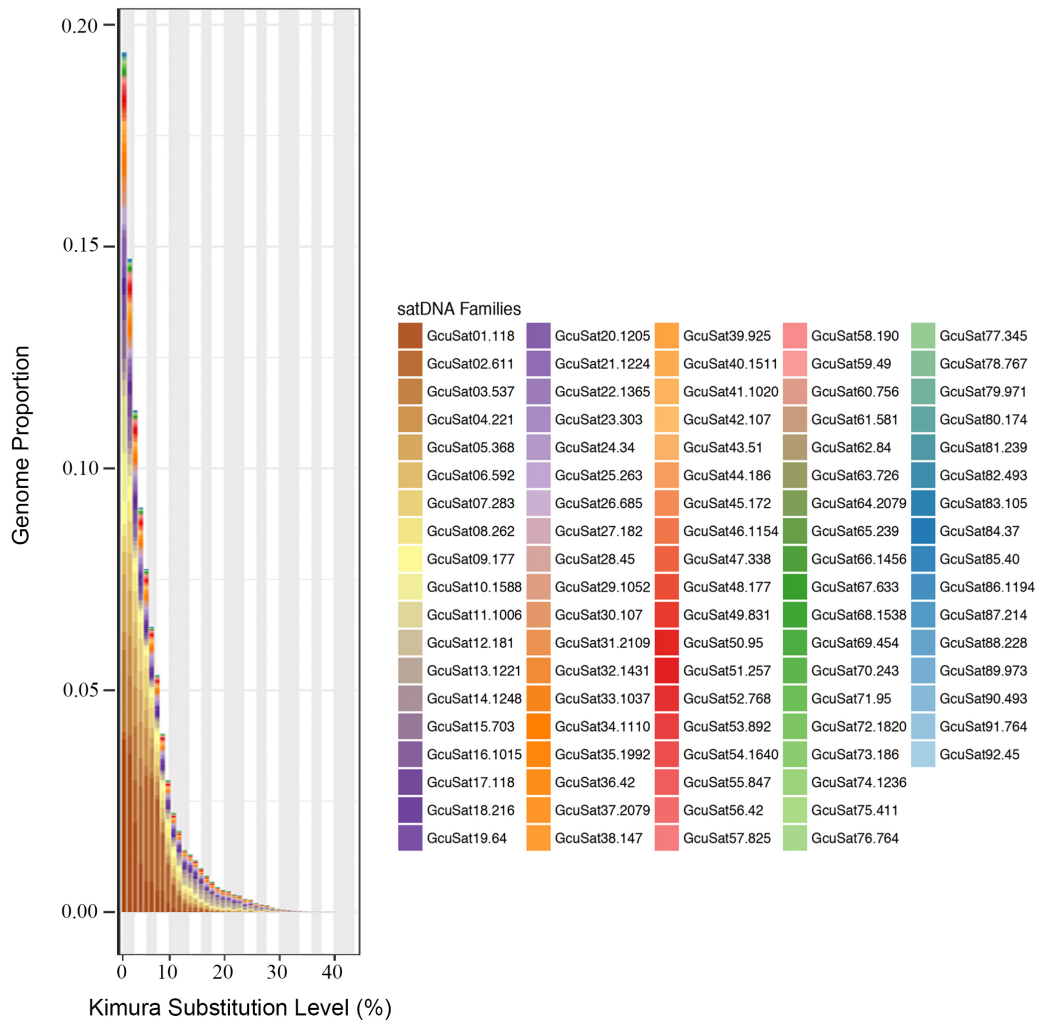


Figura 1.1. Gráfico repetitivo evidenciando os perfis de abundância genômica e divergência de todos os satélites extraídos do genoma de *G. cuia*.

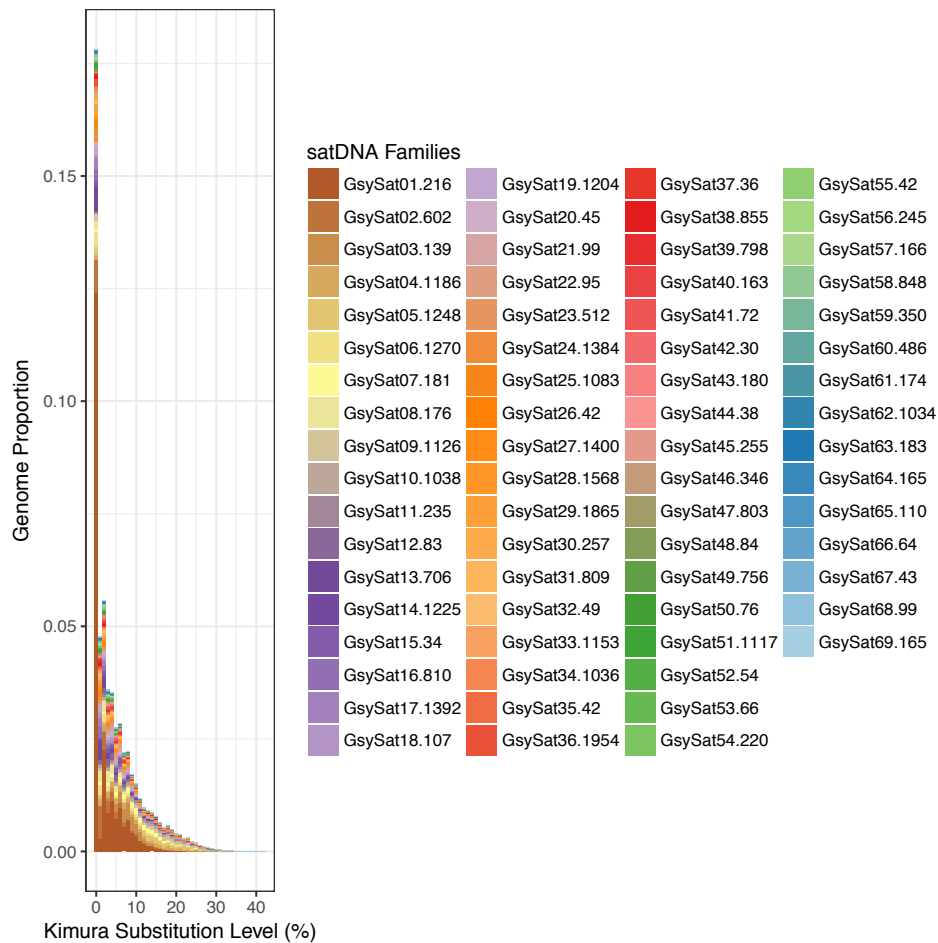


Figura 1.2. Gráfico repetitivo evidenciando os perfis de abundância genômica e divergência de todos os satélites extraídos do genoma de *G. sylvius*.

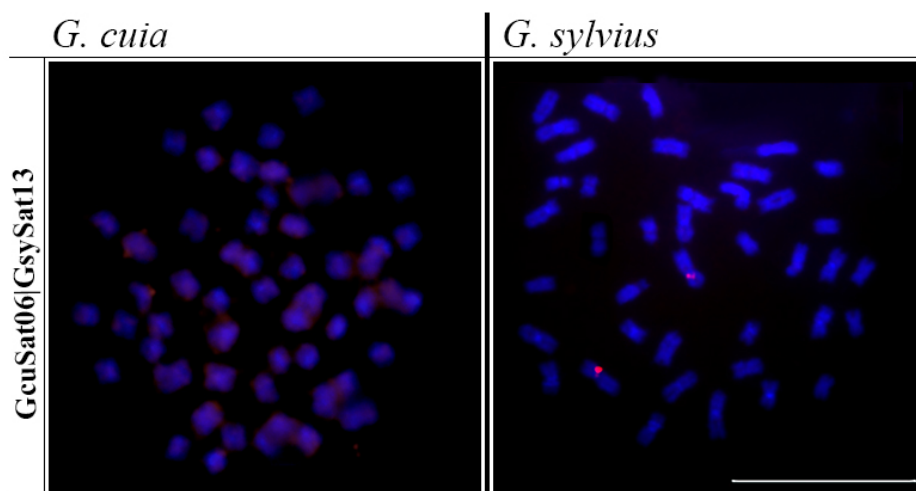


Figura 1.3. Distribuição cromossômica dos satélites GcuSat06|GsySat13 (71% de identidade de sequência) em metáfases de *G. cuia* e *G. sylvius*. Barra = 10 µm.

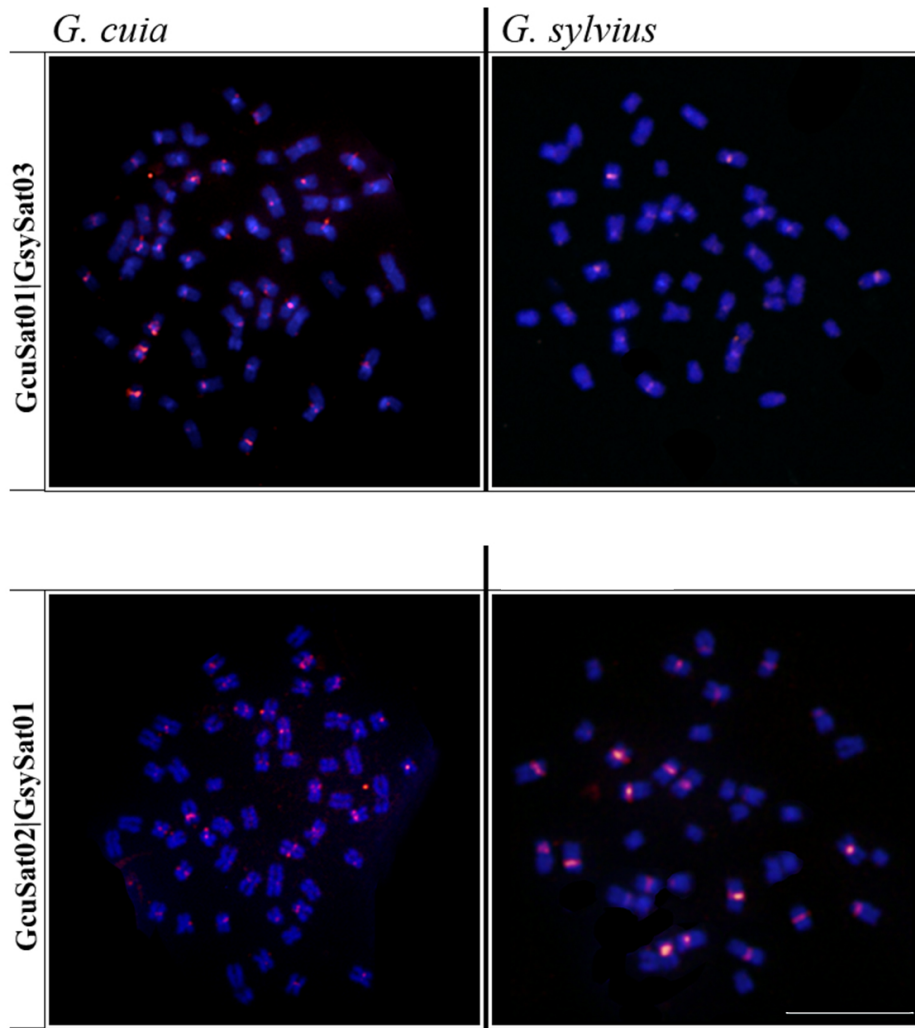


Figura 1.4. Mapeamento físico dos satélites GcuSat01|GsySat03 e GcuSat02|GsySat01, satélites com mais de 80% de identidade de sequência, nas metáfases de *G. cuia* e *G. sylvius*. Barra = 10 μ m.

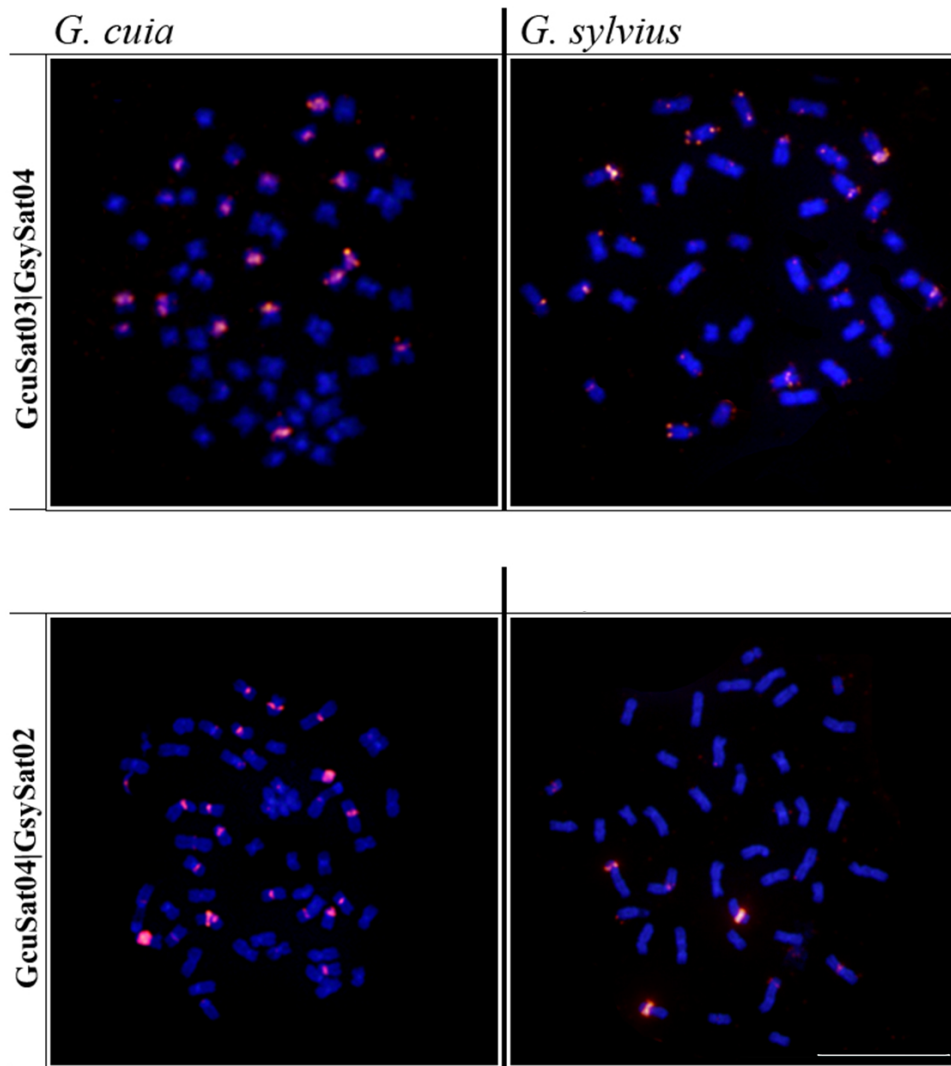


Figura 1.5. Mapeamento físico dos satélites GcuSat03|GsySat04 e GcuSat04|GsySat02, satélites com mais de 80% de identidade de sequência, nas metáfases de *G. cuia* e *G. sylvius*. Barra = 10 μ m.

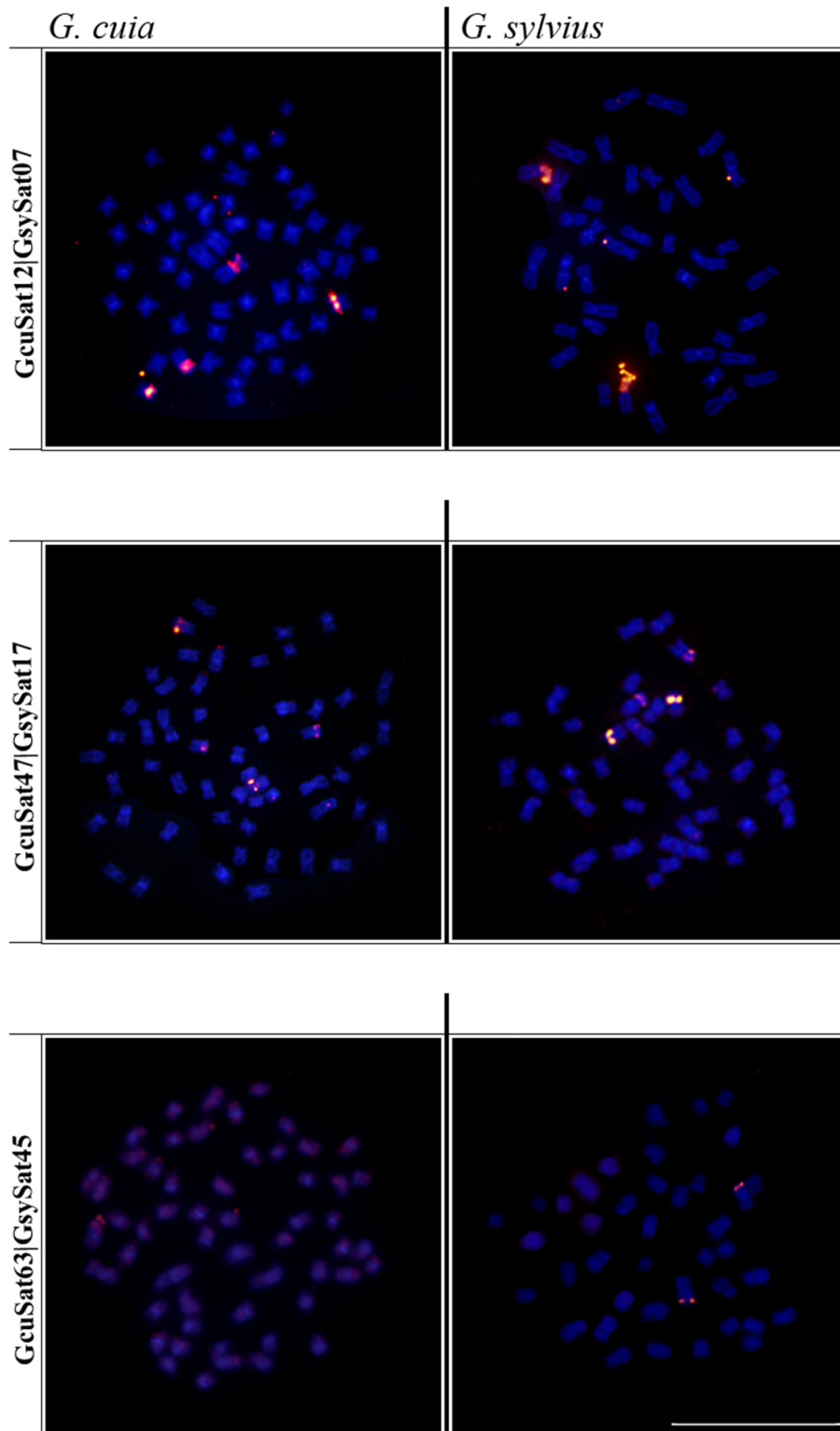


Figura 1.6. Mapeamento físico dos satélites GcuSat12|GsySat07, GcuSat47|GsySat17 e GcuSat63|GsySat45, satélites com mais de 80% de identidade de sequência, nas metáfases de *G. cuia* e *G. sylvius*. Barra = 10 μm.

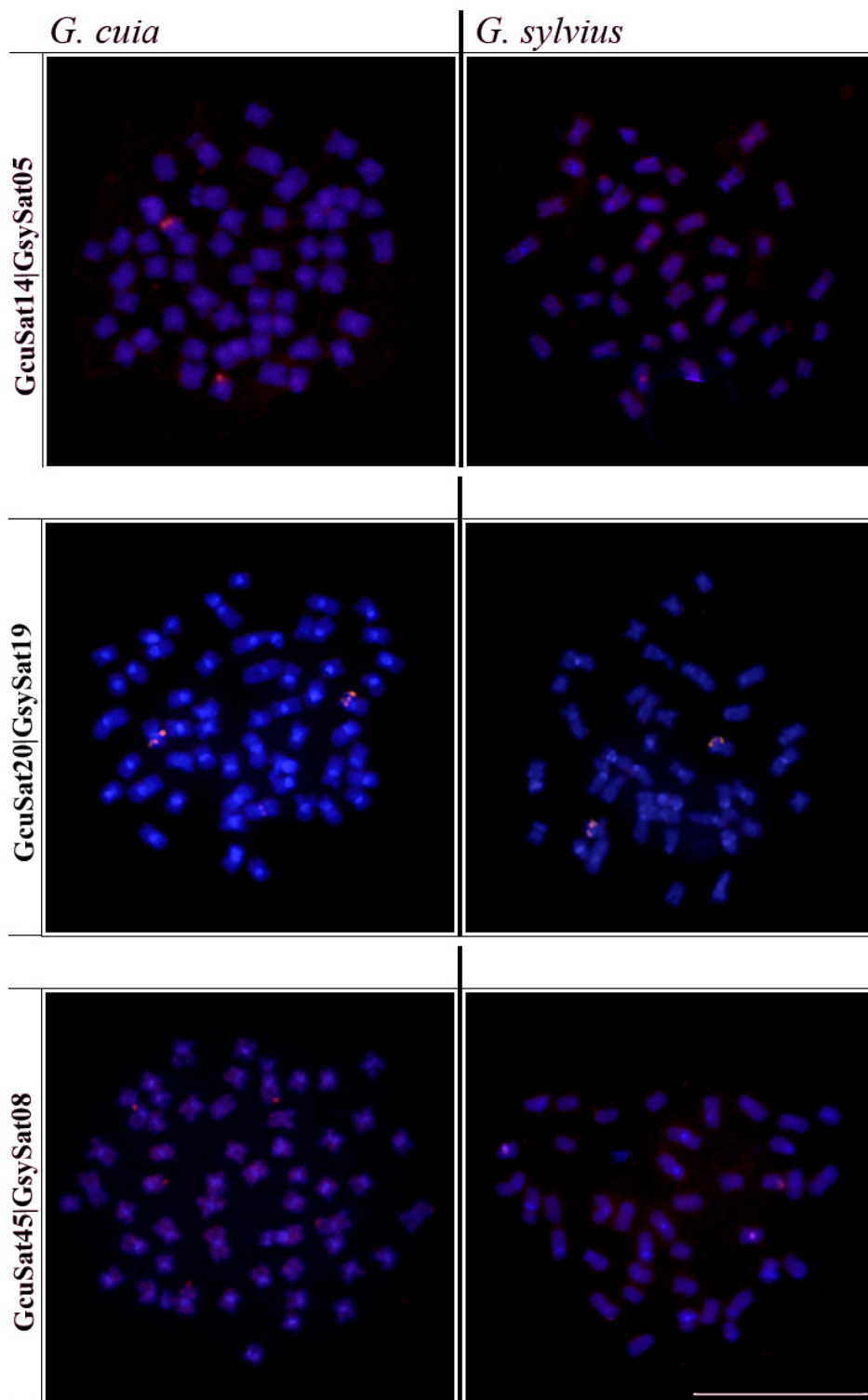


Figura 1.7. Mapeamento físico de satélites compartilhados entre *G. cuia* e *G. sylvius* que apresentam mais de 95% de identidade de sequência. Barra = 10 μm .

Discussão

Os resultados obtidos caracterizam o primeiro estudo do satelitoma em espécies de Gymnotiformes, com 92 famílias de DNAs satélites identificadas no genoma de *G. cuia* e 69 famílias extraídas no genoma de *G. sylvius*. Esse resultado contribui pro catálogo de satélites de espécies de peixes neotropicais, juntamente com os resultados obtidos em *Astyanax paranae*, *Characidium gomesi* e *Megaleporinus macrocephalus* (Silva et al. 2017; Utsunomia et al. 2019; Serrano-Freitas et al. 2020). As diferentes sequências de DNA satélite identificadas nos genomas de *G. cuia* e *G. sylvius* reforçam a diversidade desses elementos repetitivos no genoma dos peixes, com unidades de repetição de tamanhos distintos e composição nucleotídica variável. Os monômeros encontrados nas espécies apresentaram tamanhos médios distribuídos num intervalo de 34 a 2109 nt (± 625 nt) em *G. cuia* e de 30 a 1954 nt (± 511 nt) em *G. sylvius*. O conteúdo A + T mostrou-se variável, indicando uma inclinação para satélites ricos em AT, ressaltando a diversidade estrutural das sequências de DNA satélites nas espécies analisadas.

A identificação de quatro superfamílias em *G. cuia* e duas superfamílias em *G. sylvius* indica que a evolução de DNA satélites dentro dessas espécies pode ter acontecido por eventos de duplicação e diferenciação de elementos diversos e não a partir de um grupo de sequências pré-existentes, como foi identificado em *Megaleporinus macrocephalus*, espécie na qual foram encontradas 19 superfamílias (Utsunomia et al. 2019). Em contrapartida, foi detectado um alto grau de compartilhamento de satDNAs entre as espécies aqui analisadas, apresentando diferentes níveis de identidade de sequência.

De acordo com definição proposta por Ruiz-Ruano et al. (2016a), sequências satélites que compartilham similaridade acima de 50% constituem uma superfamília, enquanto define-se como família os satélites acima de 80% de semelhança e como variantes de um mesmo satDNA as sequências que apresentam acima de 95% de similaridade. Os resultados obtidos revelaram que 43 famílias de satélites de *G. cuia* apresentam algum grau de similaridade com satélites de *G. sylvius*. Estes resultados corroboram com a hipótese *library* de evolução de DNAs satélites, que é fundamentada no fato de que diferentes famílias de DNAs satélites coexistem em um genoma, constituindo uma “biblioteca” (ou *library*) permanente de sequências que podem ser amplificadas de modo independente nos genomas de espécies ou populações diferentes enquanto passam pelo processo de diferenciação. Deste modo, uma mudança rápida no perfil de distribuição, abundância e diversidade dos satDNAs deverá ser observada até mesmo em espécies próximas (Fry & Salser 1977), como é o caso de *G. cuia* e *G. sylvius*.

A nível de superfamília, o satélite GcuSat06 é similar ao GsySat13, apresentando uma abundância maior no genoma de *G. cuia* (cerca de 4%) do que em *G. sylvius* (1%). Além disso, este satélite apresenta uma distribuição clusterizada no genoma de *G. sylvius* e dispersa no genoma de *G. cuia*, demonstrando que além da acumulação genômica diferenciada, o padrão de localização e acumulação nos cromossomos das espécies também se modificou durante a diferenciação destas espécies. O elevado dinamismo saltatório dos DNAs satélites podem ser explicados pela atuação de diferentes mecanismos genéticos responsáveis pela duplicação e espalhamento destas sequências nos genomas, entre os quais a associação com elementos transponíveis, deslizamentos da DNA Polimerase durante a replicação, a ocorrência de *crossing-over* desigual, recombinação ectópica, dentre outros (Dover 1993; McMurray 1995; Hancock 1996; Cohen et al. 2010; Milani & Cabral-de-Melo 2014; Ruiz-Ruano et al. 2015, 2016a).

Dentre os satélites de *G. cuia* e *G. sylvius* que apresentam acima de 80% de similaridade, o padrão de localização na região pericentromérica dos cromossomos e a distribuição clusterizada foram as características recorrentes nos quatro satélites mais abundantes no genoma das espécies analisadas. Entretanto, no caso do satélite GcuSat03 (similar ao GsySat04), observa-se que, apesar da distribuição pericentromérica e fortemente associada com a heterocromatina, os cromossomos de *G. sylvius* apresentaram uma acumulação adicional em regiões subterminais e na região pericentromérica do par da NOR, o que não ocorre em *G. cuia*. É interessante pontuar que o satélite em questão apresenta a unidade de repetição consideravelmente longa, acima de 500 pb. Satélites maiores que 500 pb frequentemente são derivados de retrotransposons e, geralmente, ocupam apenas um *locus* na região pericentromérica, refletindo um potencial recluso em se espalhar pelo genoma (Mestrovic et al. 2015). A acumulação de Elementos Transponíveis (TEs) em regiões pericentroméricas se deve à preferência desses elementos por regiões ricas em A+T, característica de heterocromatina centromérica (Heslop-Harrison & Schwarzacher, 2011). Visto o fato de a heterocromatina ser o “ponto de encontro” de satélites e TEs, esse padrão diferencial de acumulação, no caso do satélite GsySat04, poderia ser explicado pela associação dessas diferentes classes de elementos repetitivos que, somado à plasticidade móvel de TEs, poderia levar a acumulação desse satélite em diferentes locais (Mestrovic et al. 2015).

Mesmo apresentando abundância diferencial entre os genomas, alguns satélites apresentaram o mesmo padrão de acumulação, como o caso dos satélites GcuSat47 (similar ao GsySat17) e GcuSat12 (similar ao GsySat07). As marcações similares na região intersticial do braço curto do satélite GcuSat47/GsySat17 demonstram que, mesmo sendo o 17º satélite mais

acumulado em *G. sylvius*, esse padrão de acumulação se manteve, reforçando que a acumulação desse satélite não ocorre de forma randômica e que poderia estar associado a alguma região genômica de importância entre as duas espécies. Um exemplo disso seria o satélite GcuSat12/GsySat07, que apresenta acúmulo na Região Organizadora de Nucléolo (NOR) das duas espécies, ocupando toda a região de constrição secundária do par da NOR. A conservação da associação desse satDNA à região da NOR, sugere que esse satélite poderia desempenhar alguma função estrutural nesta região, visto que o DNAr 45S apresenta um sítio bem conservado no complexo *Gymnotus carapo* (Milhomem et al. 2013).

À parte da distribuição clusterizada, que é o padrão mais frequente por se tratar de elementos repetidos *in tandem*, alguns satélites apresentaram um padrão de acumulação mais disperso. O satélite Gcu63 (similar ao GsySat45), apesar de compor 0,1% do genoma de ambas espécies, apresenta um padrão de distribuição disperso no genoma de *G. cuia* e restritamente clusterizado em *G. sylvius*, destacando uma acumulação estrutural distinta de uma mesma família de DNA satélite nessas espécies intimamente relacionadas. Em relação aos satDNAs acima de 95% de similaridade, esse padrão de distribuição difuso foi encontrado nos satélites GcuSat14 (complementar ao GsySat05) e GcuSat45 (complementar ao GsySat08), que adicionalmente apresentaram algumas marcações em pontos, padrão já evidenciado na espécie *M. macrocephalus* (Utsunomia et al. 2019). Por outro lado, o satélite GcuSat20|GsySat19 apresentou um padrão clusterizado, entretanto com distribuição restrita, nos genomas de *G. cuia* e *G. sylvius*.

Os dados apresentados integram *Gymnotus cuia* e *G. sylvius* como quarta e quinta espécie de peixes a terem o satelitoma descrito e o primeiro estudo comparativo utilizando mais de um genoma sequenciado. A anotação de várias famílias de satDNAs em diferentes organismos tem sido reportada nos últimos anos, indicando um amplo potencial estrutural e funcional destas sequências em diferentes contextos biológicos (Dalíkova et al. 2017; Lima et al. 2017; Palacios-Gimenez et al. 2017; Fu et al. 2019, Ebrahimzadegan et al. 2019). Os dados disponibilizados evidenciam a diversidade de DNA satélite nas espécies analisadas e podem contribuir em análises acerca da diversidade de DNA satélite em outras espécies relacionadas ao grupo *Gymnotus carapo*, clado mais especioso do gênero *Gymnotus*, bem como demonstram a grande relevância para fomentar futuros estudos acerca da diversidade desses elementos repetitivos em outros grupos de peixes.

Referências

As referências correspondentes a esse capítulo encontram-se organizadas no final desse documento.

Capítulo 2

Análise genômica da evolução de DNAs satélites na família Gymnotidae (Teleostei, Gymnotiformes)

Resumo

A hipótese *library* propõe que determinado grupo biológico compartilharia alguma fração na coleção de sequências satélites, possibilitando uma análise comparativa entre conjuntos de satDNA, revelando sequências deste tipo que são conservadas em diferentes espécies e que poderiam ser usadas para investigar a evolução do DNA do satélite dentro de grupos de interesse. Estudos discorrem sobre a existência de uma diversidade ampla e interessante na distribuição de DNA repetitivo no gênero de peixes elétricos *Gymnotus*; entretanto pouco se sabe sobre a dinâmica e evolução do satDNA dentro do grupo e de outras espécies fora do clado *Gymnotus carapo*. Com o objetivo de investigar a evolução da fração de DNAs satélites no gênero *Gymnotus*, os satelitomas de espécies pertencentes aos principais clados de complexos de espécies dentro de *Gymnotus* foram caracterizados e comparados. Foi possível isolar e extrair 64 famílias de DNA satélites do genoma de *G. javari*, 49 famílias do genoma de *G. inaequilabiatus* e 54 famílias do genoma de *G. cylindricus*. Além de apresentarem satDNAs em comum, na presente análise foi possível identificar satélites presentes em todos os genomas analisados, desde *G. cuia*, linhagem que divergiu mais recentemente e representante do clado *Gymnotus carapo*. A existência de DNAs satélites compartilhados corrobora a hipótese *library*, onde espécies relacionadas apresentam um conjunto compartilhado de sequências satélites. Entretanto, grande parte dos satélites compartilhados são de sequências de ordem intermediária de abundância, evidenciando uma dualidade na evolução do DNA satélite dentro do gênero; pois mesmo apresentando um grau de compartilhamento de sequências satélites, cada grupo segue seu próprio caminho evolutivo nos processos de acumulação e fixação deste tipo de sequências em seus respectivos genomas.

Introdução

Estudos taxonômicos voltados ao gênero *Gymnotus* contribuíram para o aumento do conhecimento da diversidade do grupo em 70%, com a transição de nove para 43 espécies válidas identificadas, o que refletiu em inúmeras hipóteses morfológicas e moleculares para a compreensão dos relacionamentos filogenéticos entre as espécies alocadas nesse gênero (Craig et al. 2019). Algumas espécies ainda aguardam descrição formal e possuem o *status* de subespécies, o que por vezes pode mascarar a real diversidade dentro do grupo (Lovejoy et al. 2010; Craig et al. 2017). Contudo, inúmeras tentativas cada vez mais assertivas e robustas têm sido realizadas para entender o relacionamento filogenético dentre as espécies do gênero *Gymnotus* (Lovejoy et al. 2010; Crampton et al. 2013; Tagliacollo et al. 2016, Van der Sleen & Albert 2017; Alda et al. 2019). Recentemente, Craig et al. (2019) disponibilizaram uma massiva filogenia dentro de Gymnotidae com a proposta de seis subgêneros em *Gymnotus*, abrangendo todas as 43 espécies atualmente válidas.

A diversidade da família Gymnotidae está amplamente concentrada no gênero *Gymnotus*, visto que *Electrophorus* é o grupo menos representativo, com três espécies válidas (Santana et al. 2019). O gênero *Gymnotus* está dividido em seis complexos de espécies, de acordo com sinapomorfias taxonômicas e regiões biogeográficas de distribuição (Tagliacollo et al. 2016), compreendendo *G. pantherinus*, *G. coatesi*, *G. anguillaris*, *G. tigre*, *G. cylindricus* e *G. carapo*. As contribuições citogenéticas revelaram que o gênero apresenta uma riqueza em DNA repetitivo (Milhomem et al. 2013, Scacchetti et al. 2011, 2012; Utsunomia et al. 2014, 2018), especialmente de microssatélites. Notavelmente, o DNA ribossômico 5S apresenta diferenças significativas entre algumas espécies do complexo *G. carapo*, sendo proposto como um possível marcador biogeográfico de espécies desse grupo, ressaltando o potencial de estudos com elementos repetitivos para investigar padrões evolutivos no genoma, correlacionando-os à história natural dessas espécies (Silva et al. 2019).

Uma outra classe de DNA repetitivo, o DNA satélite (satDNA), é formada por elementos repetidos *in tandem* que apresentam alta taxa evolutiva e muitas vezes são espécie-específicos (Garrido-Ramos 2017; Novák et al. 2017), constituindo-se como potenciais marcadores citogenéticos para uma determinada espécie. Entretanto, de acordo com a hipótese *library* (Fry & Salser 1977), mesmo que cada família de satDNA sofra eventos de acumulação e homogeneização isolados dentro de cada espécie, um determinado número de satélites pode ser compartilhado pelos componentes de grupos relacionados. Neste contexto, a utilização de metodologias recentes como o Sequenciamento de Próxima Geração, em associação com

ferramentas computacionais, tem permitido um acesso mais refinado aos genomas, investigando a diversidade de DNA satélite e revelando a trajetória evolutiva desses elementos nas espécies analisadas. Dessa forma, o objetivo do presente trabalho consistiu em caracterizar o satelitoma de representantes adicionais do gênero *Gymnotus* e verificar o grau de compartilhamento de sequências de satDNAs nos genomas destas espécies de peixes elétricos neotropicais. Considera-se que as informações obtidas serão úteis para uma melhor compreensão dos processos envolvidos na diversificação e relação entre as espécies, bem como dos processos envolvidos na estruturação do genoma neste grupo de peixes.

Metodologia

As análises moleculares com espécies de clados de *Gymnotus* foram acessadas a partir de amostras da coleção ictiológica do laboratório do professor Nathan Lovejoy, da *University of Toronto*, em parceria com o *Royal Ontario Museum*, com espécies coletadas em diferentes pontos de distribuição ao longo da região Neotropical (Figura 2.1). Assim, aproximadamente 1 Gb do genoma de *G. javari* (clado *G. coatesi*), *G. inaequilabiatus* (clado *G. tigre*) e *G. cylindricus* (clado *G. cylindricus*) foram sequenciados na plataforma Illumina Miseq (*paired-end* 2x 150). Posteriormente, os dados gerados foram comparados com as famílias de DNAs satélites isolados na espécie *G. cuia*, do clado *G. carapo* (capítulo 1).

A filtragem sucessiva de elementos repetitivos foi realizada com o uso da plataforma RepeatExplorer através de arquivos gerados por *scripts* disponibilizados em Ruiz-Ruano et al. (2016a), com a checagem de todos os gráficos circulares, característico de DNA satélite, para o isolamento de sequências com auxílio do software Geneious Prime 2020. As iterações subsequentes foram realizadas até não ser possível encontrar mais nenhuma sequência de DNA satélite. Cada biblioteca foi analisada separadamente para a classificação dos satélites em ordem decrescente de abundância e os satelitomas entre as espécies foram comparados através do alinhamento *De Novo*, disponibilizado no software Geneious Prime 2020. Uma comparação excludente dos satelitomas, a partir do genoma de *G. cuia*, foi realizada com a finalidade de detectar possíveis satélites ancestrais no gênero, presentes nos grupos analisados, respeitando a filogenia mais recente e aceita dentro do grupo (Tagliacollo et al. 2016; Craig et al. 2019).

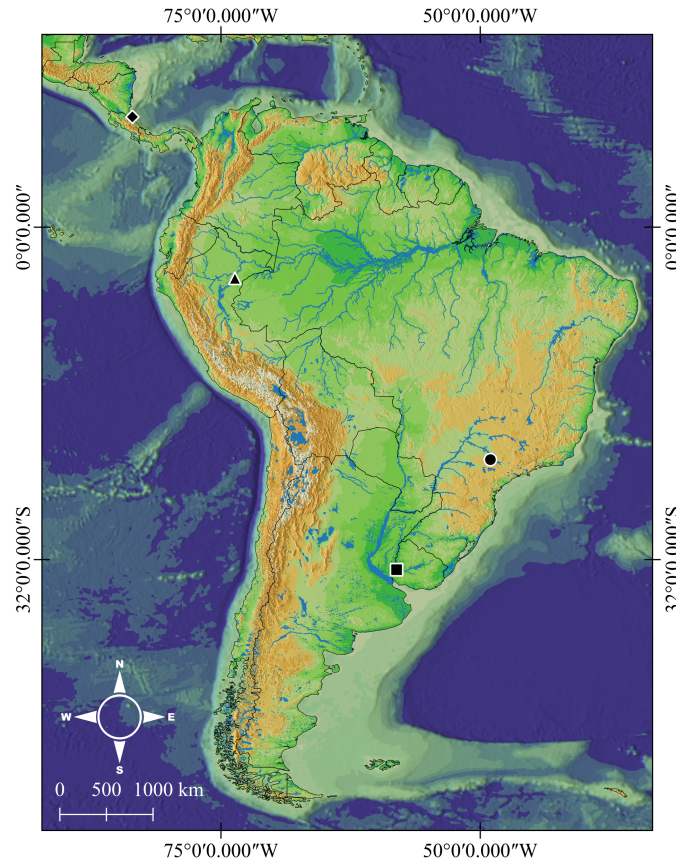


Figura 2.1. Localidades de amostragem das espécies *G. javari* (triângulo), *G. inaequilabiatus* (quadrado), *G. cylindricus* (polígono) e *G. cuiá* (círculo).

Resultados

O sequenciamento e análise da composição do genoma dessas três espécies de peixes da família Gymnotidae permitiram identificar 64 famílias de DNAs satélites em *Gymnotus javari* (clado *G. coatesi*), 49 famílias em *G. inaequilabiatus* (clado *G. tigre*) e 54 famílias em *G. cylindricus* (Figura 2.2, 2.3 e 2.4). Informações específicas de cada família de DNA satélite isolada considerando parâmetros como tamanho, abundância e divergência de monômeros nas sequências estão disponíveis nos materiais suplementares (Tabelas Suplementares 2.1, 2.2 e 2.3). Os dados sobre a diversidade geral de tamanho de monômero, porcentagem de A+T e total de sequências satélites, com a adição das variantes das espécies analisadas estão relacionados na Tabela 2.1.

Tomando por base as 43 famílias de satDNAs compartilhadas entre *G. cuiá* e *G. sylvius* (analisadas no capítulo anterior) como ponto de partida da análise excludente, comparando sucessivamente as linhagens mais recentes até as mais antigas, observou-se uma diminuição progressiva no número de satélites compartilhados até ser possível identificar quatro sequências

satélites que estão presentes em todas as bibliotecas filtradas: GcuSat21, GcuSat46, GcuSat52 e GcuSat55 (Figura 2.5 e 2.6). Esses satélites, denominados ancestrais, estão presentes em todas as espécies analisadas, com variados níveis de abundância e divergência ao longo dos genomas (Tabela 2.2). Adicionalmente, a análise comparativa dos satelitomas após o alinhamento *De Novo* revelou um número variado de satélites compartilhados entre os grupos (Tabela 2.3).

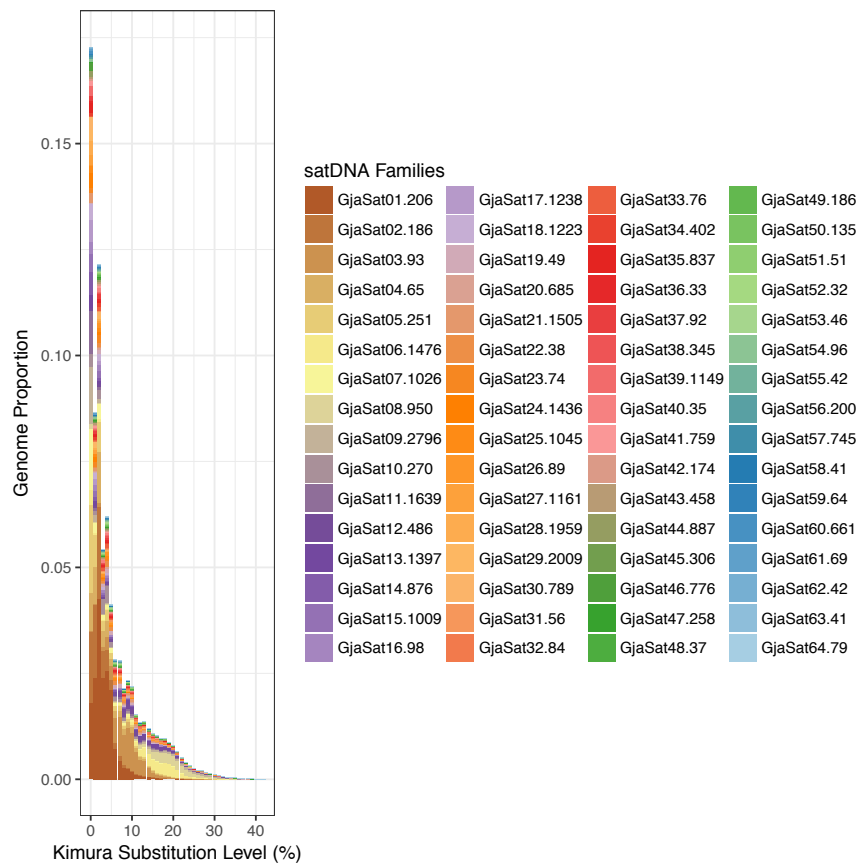


Figura 2.2. Gráfico com o panorama de seqüências satélites, correlacionando os perfis de abundância e divergência genômica do satelitoma extraído de *G. javari*.

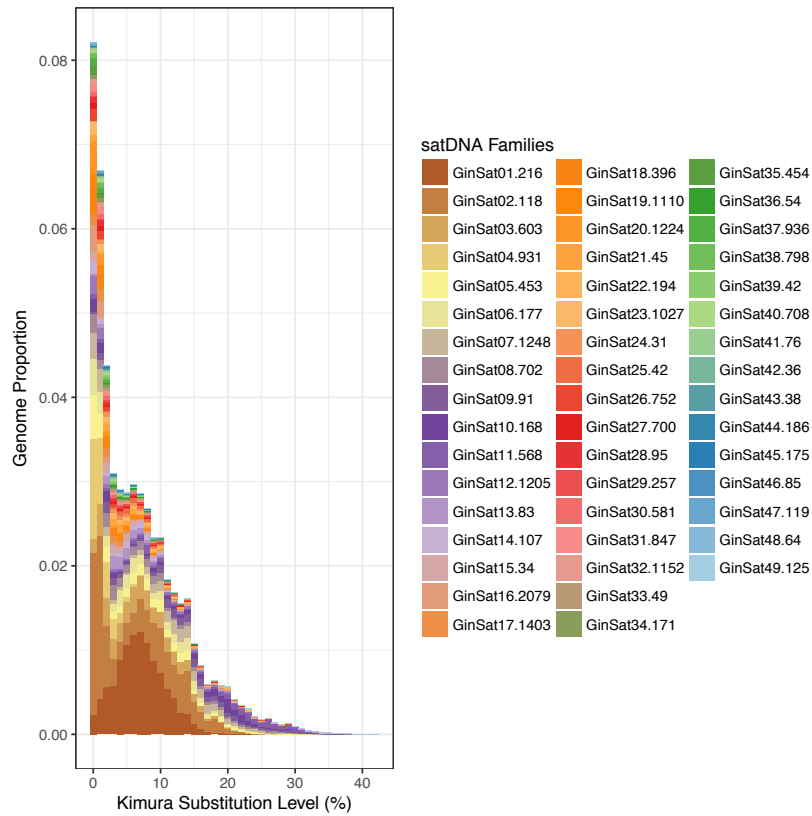


Figura 2.3. Gráfico com o panorama de seqüências satélites, correlacionando os perfis de abundância e divergência genômicas do satelitoma extraído de *G. inaequilabiatus*.

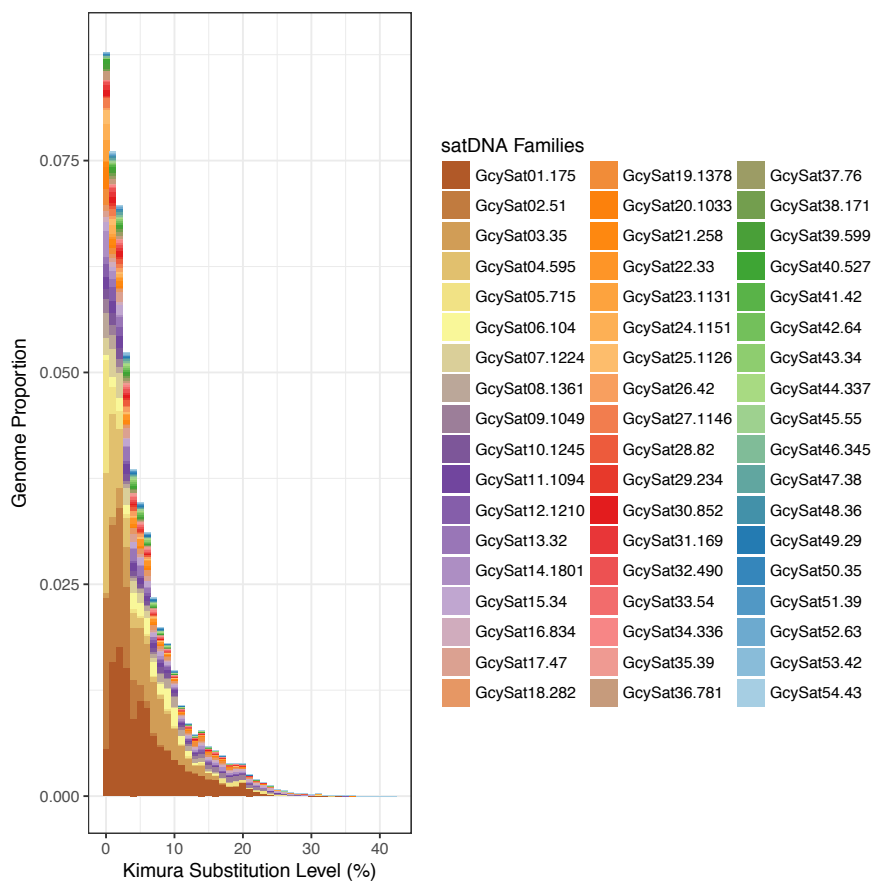


Figura 2.4. Gráfico com o panorama de seqüências satélites, correlacionando os perfis de abundância e divergência genômica do satelitoma extraído de *G. cylindricus*.

Tabela 2.1. Diversidade geral de DNA satélites em *G. javari*, *G. inaequilabiatus* e *G. cylindricus*. Os cálculos de média e porcentagem A+T foram realizados com a variante mais abundante dentro de cada família.

satDNA	Menor monômero (nt)	Maior monômero (nt)	Média (nt)	Média A + T (%)	Total de satDNAs com variantes
<i>G. javari</i> (GjaSat)	32	2796	553	57,1	72
<i>G. inaequilabiatus</i> (GinSat)	31	2079	464	57,2	63
<i>G. cylindricus</i> (GcySat)	29	1801	459	57,4	69

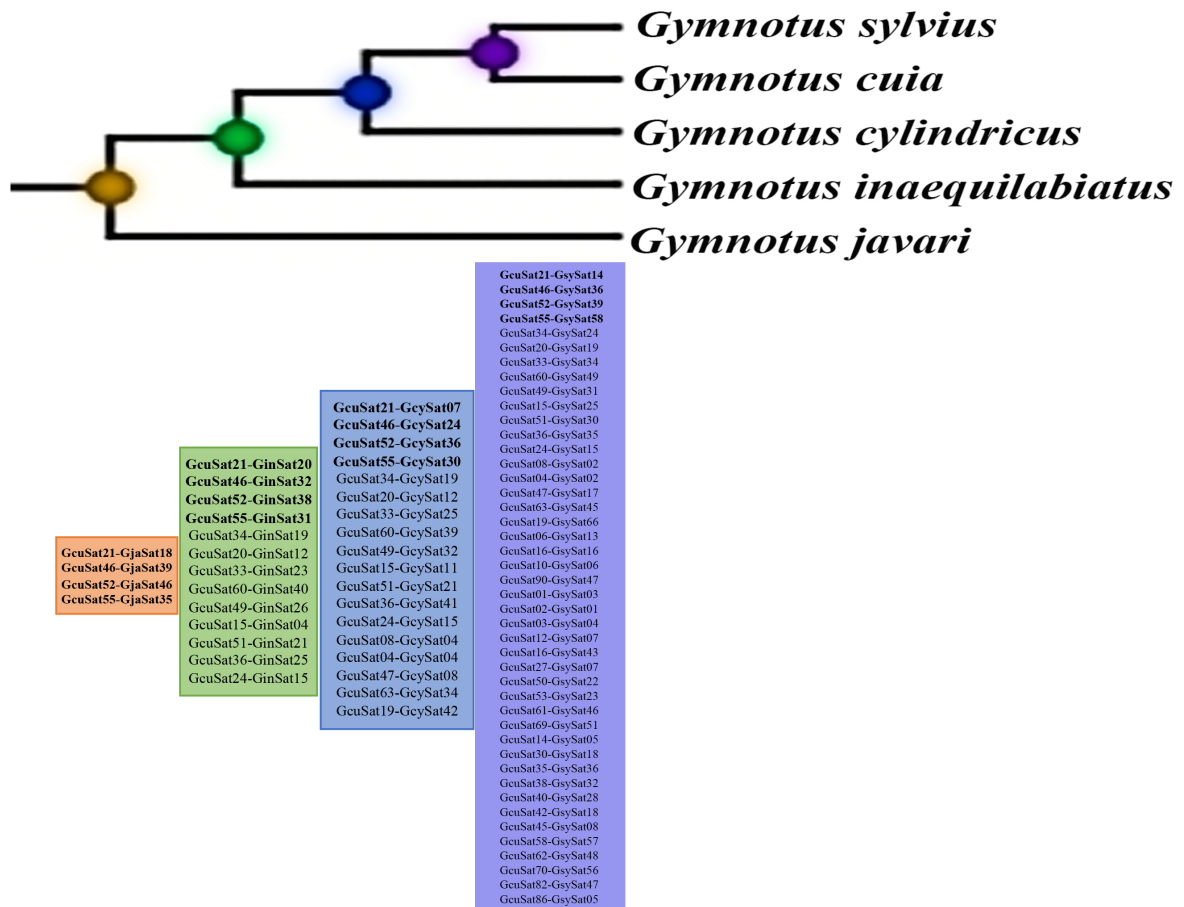


Figura 2.5. Satélites ancestrais em *Gymnotus*. O grau de compartilhamento abrange desde todas as espécies analisadas (laranja), entre *G. cuia* + *G. sylvius*, *G. cylindricus* e *G. inaequilabiatus* (verde), *G. cuia* + *G. sylvius* e *G. cylindricus* (azul) e *G. cuia* + *G. sylvius* (roxo). Os quatro satélites ancestrais a todas as espécies (laranja) apresentaram uma identidade de sequência acima de 80%. (Topologia modificada de Alda et al. 2019).

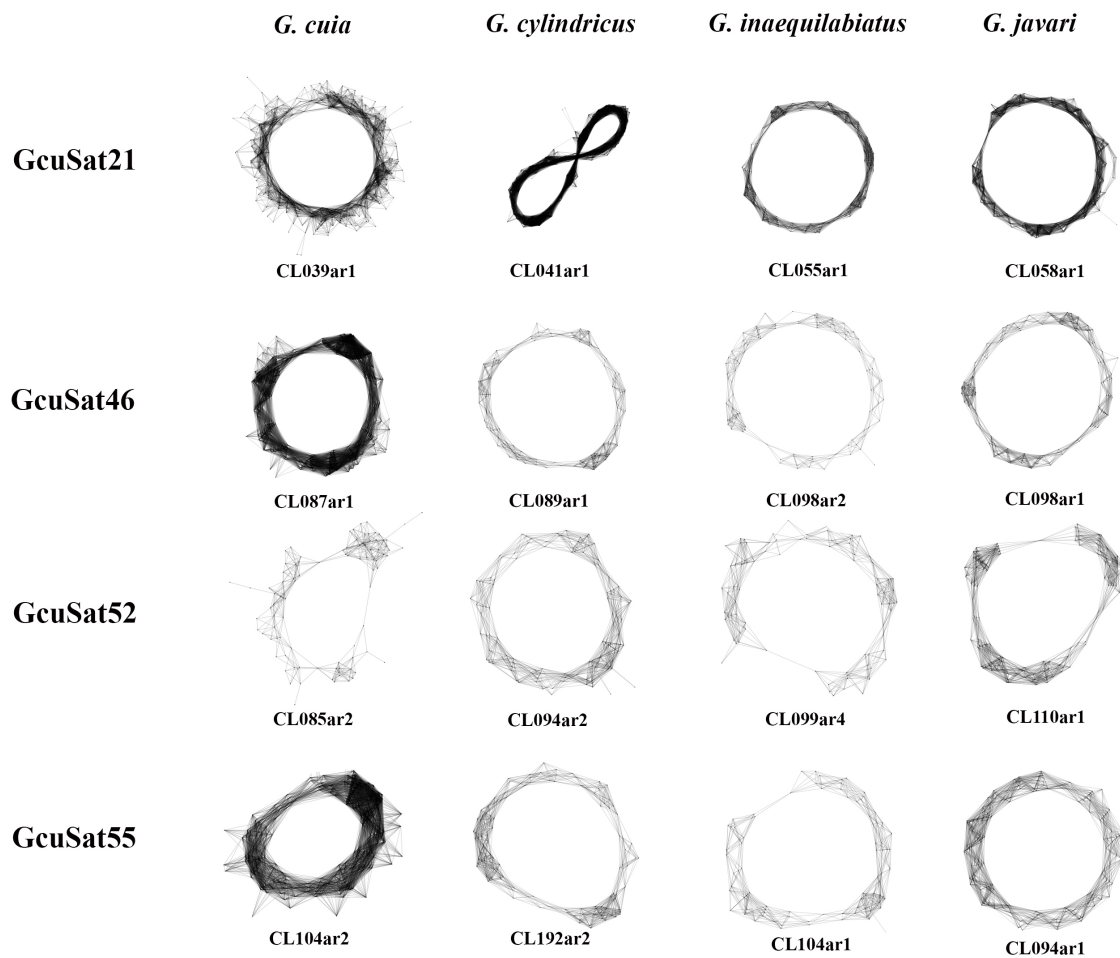


Figura 2.6. Representação gráfica dos clusters que contêm os satélites ancestrais isolados de *G. cuia*, *G. cylindricus*, *G. inaequilabiatus* e *G. javari*.

Tabela 2.2. Relação dos quatro satélites isolados de *G. cuia* (capítulo 1) que estão presentes no genoma de todas as espécies analisadas, com o satélite correspondente em cada espécie, abundância e divergência genômicas.

Sats Ancestrais	GcuSat21-1224	GcuSat46-1154	GcuSat52-768	GcuSat55-847
Abundância em <i>G. cuia</i>	0,009458	0,002951	0,002356	0,001939
Divergência em <i>G. cuia</i>	0,74	2,34	1,4	1,61
Correspondência em <i>G. cylindricus</i>	GcySat07-1224	GcySat24-1151	GcySat36-781	GcySat30-852
Abundância em <i>G. cylindricus</i>	0,014673	0,003632	0,002206	0,002861
Divergência em <i>G. cylindricus</i> (%)	2,85	2,26	1,31	2,66
Correspondência em <i>G. inaequilabiatus</i>	GinSat20-1224	GinSat32-1152	GinSat38-798	GinSat31-847
Abundância em <i>G. inaequilabiatus</i>	0,006206	0,002349	0,001716	0,002412
Divergência em <i>G. inaequilabiatus</i> (%)	0,69	4,1	1,44	1,49
Correspondência em <i>G. javari</i>	GjaSat18-1223	GjaSat39-1149	GjaSat46-776	GjaSat35-837
Abundância em <i>G. javari</i>	0,008346	0,003368	0,002458	0,004145

Divergência em <i>G. javari</i> (%)	1.76	4.52	0.99	3.49
-------------------------------------	------	------	------	------

Tabela 2.3. Satélites compartilhados entre as diferentes bibliotecas de *Gymnotus* analisadas.

GcuSat GcySat	GcySat GinSat	GcuSat GinSat	GinSat GjaSat	GcySat GjaSat	GcuSat GjaSat
GcuSat04 GcySat04	GcySat05 GinSat08	GcuSat01 GinSat02	GinSat01 GjaSat01	GcySat07 GjaSat18	GcuSat02 GjaSat02
GcuSat08 GcySat04	GcySat07 GinSat20	GcuSat02 GinSat01	GinSat01 GjaSat02	GcySat15 GjaSat08	GcuSat21 GjaSat18
GcuSat15 GcySat11	GcySat08 GinSat48	GcuSat03 GinSat06	GinSat04 GjaSat25	GcySat19 GjaSat21	GcuSat46 GjaSat39
GcuSat19 GcySat42	GcySat12 GinSat12	GcuSat06 GinSat03	GinSat20 GjaSat18	GcySat24 GjaSat39	GcuSat47 GjaSat17
GcuSat20 GcySat12	GcySat15 GinSat15	GcuSat09 GinSat06	GinSat22 GjaSat43	GcySat26 GjaSat55	GcuSat52 GjaSat46
GcuSat21 GcySat07	GcySat17 GinSat21	GcuSat10 GinSat08	GinSat31 GjaSat35	GcySat30 GjaSat35	GcuSat55 GjaSat35
GcuSat22 GcySat27	GcySat19 GinSat19	GcuSat11 GinSat05	GinSat32 GjaSat39	GcySat33 GjaSat31	GcuSat56 GjaSat55
GcuSat24 GcySat15	GcySat21 GinSat29	GcuSat14 GinSat07	GinSat33 GjaSat14	GcySat34 GjaSat47	GcuSat59 GjaSat19
GcuSat25 GcySat29	GcySat23 GinSat35	GcuSat15 GinSat04	GinSat33 GjaSat19	GcySat36 GjaSat46	GcuSat61 GjaSat34
GcuSat28 GcySat17	GcySat24 GinSat32	GcuSat16 GinSat18	GinSat35 GjaSat30	GcySat37 GjaSat33	GcuSat62 GjaSat32
GcuSat32 GcySat09	GcySat25 GinSat23	GcuSat17 GinSat05	GinSat36 GjaSat31	GcySat38 GjaSat49	GcuSat63 GjaSat47
GcuSat33 GcySat25	GcySat26 GinSat32	GcuSat18 GinSat05	GinSat38 GjaSat46	GcySat46 GjaSat38	GcuSat69 GjaSat30
GcuSat34 GcySat19	GcySat26 GinSat39	GcuSat20 GinSat12	GinSat39 GjaSat55	12 contigs	GcuSat73 GjaSat49
GcuSat36 GcySat41	GcySat26 GinSat43	GcuSat21 GinSat20	GinSat40 GjaSat24		GcuSat77 GjaSat38
GcuSat41 GcySat09	GcySat27 GinSat17	GcuSat22 GinSat17	GinSat41 GjaSat33		GcuSat89 GjaSat56
GcuSat46 GcySat24	GcySat29 GinSat22	GcuSat24 GinSat15	GinSat43 GjaSat22		15 contigs
GcuSat47 GcySat08	GcySat30 GinSat31	GcuSat27 GinSat09	GinSat44 GjaSat49		
GcuSat49 GcySat32	GcySat33 GinSat36	GcuSat27 GinSat53	17 contigs		
GcuSat51 GcySat21	GcySat36 GinSat38	GcuSat28 GinSat21			
GcuSat52 GcySat36	GcySat37 GinSat41	GcuSat30 GinSat14			
GcuSat55 GcySat30	GcySat38 GinSat44	GcuSat33 GinSat23			
GcuSat56 GcySat26	GcySat41 GinSat25	GcuSat34 GinSat19			
GcuSat57 GcySat16	GcySat42 GinSat10	GcuSat36 GinSat25			
GcuSat60 GcySat39	GcySat42 GinSat48	GcuSat37 GinSat16			
GcuSat63 GcySat34	24 contigs	GcuSat39 GinSat37			
GcuSat73 GcySat38		GcuSat42 GinSat14			
GcuSat77 GcySat46		GcuSat45 GinSat34			
GcuSat92 GcySat17		GcuSat46 GinSat32			
28 contigs		GcuSat49 GinSat26			
		GcuSat50 GinSat28			
		GcuSat51 GinSat29			
		GcuSat52 GinSat38			
		GcuSat55 GinSat31			
		GcuSat56 GinSat39			
		GcuSat58 GinSat47			

GcuSat59|GinSat33
GcuSat60|GinSat40
GcuSat61|GinSat30
GcuSat64|GinSat16
GcuSat69|GinSat35
GcuSat71|GinSat09
GcuSat73|GinSat44
GcuSat80|GinSat45

43 contigs

Discussão

O conjunto de satDNAs satélites aqui descritos revelou uma elevada diversidade e abundância destes elementos nos componentes do gênero *Gymnotus* e contribui para a anotação de novos catálogos de satDNAs em peixes neotropicais. Destaca-se que foi possível identificar um conjunto de quatro satélites ancestrais, compartilhados por todas as espécies dos diferentes clados, corroborando a hipótese *library* (Fry & Salser 1977; Ruiz-Ruano et al. 2016a; Garrido-Ramos 2017). Estes satélites compartilhados apresentaram uma abundância e divergência baixas, indicando que essas sequências estão homogeneizadas nas bibliotecas analisadas e podem ter divergido, de fato, a mais tempo. Entretanto, em *G. cylindricus*, houve uma acumulação do satélite correspondente ao GcuSat21, o que pode ser justificado por eventos únicos de amplificação dessa sequência devido a mecanismos de evolução em concerto do DNA satélite (Mestrovic et al. 2006).

As análises comparativas dos satelitomas revelaram uma diversidade de elementos satélites compartilhados que possibilitam investigar a evolução do DNA satélite dentro do gênero *Gymnotus*. O compartilhamento mais restrito de sequências satélites entre *G. cuiia* e *G. javari* é justificado por essas espécies terem divergido a mais tempo, sendo filogeneticamente mais distantes. Entretanto, é interessante destacar o compartilhamento de alguns satélites com maior abundância entre essas espécies, no caso da relação do GcuSat02 com GjaSat02, o segundo satélite mais abundante em *G. javari*, responsável por constituir aproximadamente 7% da biblioteca genômica da espécie.

Os dois primeiros satélites em compartilhamento de todas as espécies analisadas compõem, juntos, cerca de 20% da biblioteca genômica filtrada (vide tabelas suplementares do

capítulo 2). Foi possível verificar que esses satélites possuem algum nível de identidade de sequência no comparativo geral de satelitomas, como é o caso do GcuSat01 com GinSat02, GcuSat02 com GinSat01, GinSat01 com GjaSat01, GinSat01 com GjaSat02 e GcuSat02 com GjaSat02. Curiosamente, os satélites mais abundantes de *G. cylindricus* que tiveram alguma relação com satélites de outros genomas foram GcySat04 (complementar ao GcuSat04) e GcySat05 (complementar ao GinSat08) que constituem 5% e 4% do genoma, respectivamente. Essa disparidade encontrada no satelitoma comparativo poderia ser decorrente do fato de *G. cylindricus* estar isolado das outras espécies. Tal situação atuaria de modo a fixar e homogeneizar sequências satélites que se modificam a partir de eventos estocásticos de amplificação pontual, devido à atuação de mecanismos responsáveis pela modificação de sequências satélites, como a associação com elementos transponíveis, deslizamentos da DNA polimerase durante a replicação e *crossing-over* desigual (Ruiz-Ruano et al. 2016b; Garrido-Ramos 2017).

No genoma das espécies de *Gymnotus carapo* (Capítulo 1) observou-se que há um grande número de sequências de DNA satélites abundantes compartilhado entre *G. cuia* e *G. sylvius*, sugerindo que em espécies proximamente relacionadas o número de satélites de maior ordem de abundância compartilhados é maior do que em espécies mais distantes. Grupos filogeneticamente mais distantes possuem uma maior diferenciação genética, como verificado ocorrer nos gafanhotos dos gêneros *Locusta* e *Eumigus*, que não compartilham nenhuma família de DNA satélite entre si, sugerindo que a divergência entre esses dois gêneros é um evento muito antigo (Ruiz-Ruano et al. 2016b). No caso das espécies de *Gymnotus* analisadas, é possível constatar que há uma porção de biblioteca genômica compartilhada, mas entre satélites de ordem intermediária de abundância genômica, indicando que cada grupo apresenta o seu próprio caminho evolutivo no que tange à acumulação e dispersão de DNAs satélites.

A análise comparativa revelou que há um compartilhamento de maior número de satélites entre *G. inaequilabiatus* e *G. cuia* do que entre *G. cylindricus* e *G. cuia*, considerados mais próximos filogeneticamente (clado *G. carapo*), de acordo com filogenias mais recentes (Lovejoy et al. 2010; Tagliacollo et al. 2016, Alda et al. 2019, Craig et al. 2019). Essa relação maior entre GinSats e GcuSats poderia ser explicada pelo fato de que a proximidade espacial entre as espécies poderia refletir em condições similares de geodispersão e pressões evolutivas, constituindo-se no principal fator para a homogeneização e posterior fixação de sequências satélites entre os grupos.

A plasticidade cromossômica no gênero *Gymnotus* é comumente associada a eventos de rearranjo e associações de sequências satélites, que são elementos bastante dinâmicos (Silva et al. 2019). Os resultados obtidos no presente estudo contribuem para o conhecimento da dinâmica evolutiva das sequências de DNA satélite dentro do gênero *Gymnotus*, podendo, também, subsidiar futuras análises com marcadores citogenéticos para investigar a localização de satélites compartilhados entre diferentes grupos, especialmente dos clados *G. tigre* e *G. cylindricus*, sobre os quais ainda não se tem informações acerca de números diploides e mapeamento físico de sequências repetitivas.

Referências

As referências correspondentes a esse capítulo encontram-se organizadas no final desse documento.

5 CONSIDERAÇÕES FINAIS

Segmentos de DNA repetitivo, anteriormente associados à depósitos genômicos inertes e sem nenhuma função aparente, recentemente têm estado sob a atenção de pesquisas por exibirem mais do que apenas funções estruturais. Dentre as novas funções que estão sendo atribuídas a esses elementos do genoma, ressaltam-se aquelas relacionadas ao enovelamento do material genético e à regulação de genes, caso específico do DNA satélite, uma das mais diversas classes de DNA repetitivo no genoma. Os resultados apresentados evidenciam a riqueza de DNA satélites em *Gymnotus* e podem contribuir para estudos futuros envolvendo a montagem do genoma deste grupo de peixes, visto que a relação de elementos repetitivos e sua abundância geralmente representam um desafio na montagem de genomas, devido a sua dinâmica própria e ao alto grau de polimorfismo dessas sequências.

Analisando estruturalmente os satDNAs identificados em *G. cuia* que apresentaram uma composição nucleotídica altamente similar com os satélites de *G. sylvius*, é possível afirmar que, embora identificadas como espécies distintas, o satelitoma dessas espécies é bem conservado, o que pode ser explicado pelo fato de as duas espécies pertencerem ao grupo *Gymnotus carapo* e, portanto, apresentarem mais proximidade nas suas relações filogenéticas. Em contrapartida, é possível destacar a existência de uma abundância variável na acumulação genômica e estrutural dentre esses elementos, corroborando a definição canônica de constituírem sequências repetitivas *in tandem*, com a organização altamente variável por desenvolverem caminhos evolutivos independentes.

Sequências satélites em distintos estágios de evolução e acumulação no genoma das espécies demonstram que, mesmo possuindo uma biblioteca ancestral, as espécies sofrem diferentes pressões evolutivas que refletem na fixação e acumulação genômica desses elementos, o que pode ser reforçado pela ausência de comportamento migratório no grupo e consequente isolamento geográfico dessas espécies (Silva et al. 2019). Ao realizar uma comparação excludente de satélites ao longo dos grupos, foi possível identificar um conjunto ancestral de satDNAs presentes em todas as espécies analisadas. Entretanto, os satélites compartilhados são de ordem intermediária de abundância, indicando que cada espécie tem um processo evolutivo independente na acumulação e fixação de DNAs satélites mais abundantes.

A caracterização de DNAs satélites em estudos anteriores foi realizada, por vezes, através de abordagens que permitiam o isolamento moderado de sequências, o que muitas vezes mascarava a riqueza de satDNAs existentes em determinados grupos (Garrido-Ramos 2017).

Além disso, por estarem comumente associados à heterocromatina, erroneamente considerada “DNA-lixo”, achava-se que essas sequências também não apresentavam função (Garrido-Ramos, 2017). Com a incorporação de novas metodologias potencialmente resolutivas, como NGS, análises *in silico* e ferramentas de citogenética molecular, o estudo do DNA satélite está se tornando uma abordagem eficaz e promissora no estudo da evolução cariotípica e genômica de variados grupos biológicos.

6 REFERÊNCIAS

- Albert, J. S. Species diversity and phylogenetic systematics of American knifefishes (Gymnotiformes, Teleostei). Los Angeles: Museum of Zoology, *University of Michigan* 127 (2001).
- Albert, J. S., Crampton, W. G. R. Seven new species of the Neotropical electric fish *Gymnotus* (Teleostei, Gymnotiformes) with a redescription of *G. carapo* (Linnaeus). *Zootaxa* 1-54, (2003).
- Alda, F., Tagliacollo, V. A. Bernt, M. J., Waltz, B. T., Ludt, W. B., Faircloth, B. C., Alfaro, M. E., Albert, J. S., Chakrabarty, P. Resolving deep nodes in an ancient radiation of Neotropical fishes in the presence of conflicting signals from incomplete lineage sorting. *Syst. Biol.* **68**, 573-593 (2019).
- Arai, R. Fish karyotypes: a check list. *Tokyo: Springer Japan* 340 (2011).
- Artoni, R. F., Vicari, M. R., Endler, A. L., Cavallaro, Z. I., Jesus, C. M., Almeida, M. C., Moreira-Filho, O., Bertollo, L. A. C. Banding pattern of A and B chromosomes of *Prochilodus lineatus* (Characiformes, Prochilodontidae), with comments on B chromosomes evolution. *Genetica* **127**, 277-284 (2006).
- Britten, R. J., Graham, D. E., Neufeld, B. R. Analysis of repeating DNA Sequences by Reassociation. *Meth. Enzymol.* **29**, 363-418 (1974).
- Charlesworth, B., Snlegowski, P., Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215-220 (1994).
- Cioffi, M. B., Camacho, J. P. M., Bertollo, L. A. C. Repetitive DNAs and differentiation of sex chromosomes in Neotropical fishes. *Cytogenet. Genome Res.* **132** 188-194 (2011).
- Cohen, S., Agmon, N., Sobol, O., Segal, D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob. Dna* **1**, 11 (2010).
- Craig, J., Crampton, W., Albert, J. Revision of the polytypic electric fish *Gymnotus carapo* (Gymnotiformes, Teleostei), with descriptions of seven subspecies. *Zootaxa* **4318**, 401-438 (2017).
- Craig, J., Malabarba, L., Crampton, W., Albert, J. Revision of Banded Knifefishes of the *Gymnotus carapo* and *G. tigre* clades (Gymnotidae Gymnotiformes) from the Southern Neotropics. *Zootaxa* **4379** (2018).

- Craig, J. M., Kim, L. Y., Tagliacollo, V. A., Albert, J. S. Phylogenetic revision of Gymnotidae (Teleostei: Gymnotiformes), with descriptions of six subgenera. *PLoS ONE* **14**, e0224599 (2019).
- Crampton, W. G. R., Hopkins, C. D. Nesting and paternal care in the weakly electric fish *Gymnotus* (Gymnotiformes: Gymnotidae) with descriptions of larval and adult electric organ discharges of two species. *Copeia* **1**, 48-60 (2005).
- Crampton, W. G. R., Rodríguez-Cattáneo, A., Lovejoy, N. R., Caputi, A. A. Proximate and ultimate causes of signal diversity in the electric fish *Gymnotus*. *J. Exp. Biol.* **216**, 2523-2541 (2013).
- Dalíková, M., Zrzavá, M., Kubíčková, S., Marec, F. W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res.* **25**, 241-252 (2017).
- De La Herrán, R., Fontana, F., Lanfredi, M., Congiu, L., Leis, M., Rossi, R., Rejón, C. R., Rejón, M. R., Garrido-Ramos, M. A. Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons. *Mol. Biol. Evol.* **18**, 432-436 (2001).
- Dover, G. A. Evolution of genetic redundancy for advanced players. *Curr. Opin. Genet. Dev.*, **3**, 902-910 (1993).
- Ebrahimzadegan, R., Houben, A., Mirzaghaderi, G. Repetitive DNA landscape in essential A and supernumerary B chromosomes of *Festuca pratensis* Huds. *Sci. Rep.* **9**, 19989 (2019).
- Faria-Pereira, L. P., Hilsdorf, A. S., Albert, J., Paiva, M. J. T. R., Galvão, M. S. N. Molecular assessment of *Gymnotus* spp. (Gymnotiformes: Gymnotidae) fishing used as live baitfish in the Tietê River, Brazil. *Neotrop. Ichthyol.* **17**, e190075 (2019).
- Fernandes-Matioli, F. M. C., Almeida-Toledo, L. F. A Molecular phylogenetic analysis in *Gymnotus* species (Pisces: Gymnotiformes) with inferences on chromosome evolution. *Caryologia* **54**, 23-30 (2001).
- Foresti, F., Toledo-Filho, S. A., Almeida-Toledo, L. F. Polymorphic nature of the Nucleolus Organizer Regions in fishes. *Cytogenet. Cell Genet.* **31**, 134-141 (1981).
- Fry, K., Salser, W. Nucleotide Sequences of HS-A satellite DNA from kangaroo rat *Dipodomys Ordii* and characterization of similar sequences in other rodents. *Cell* **12**, 1069-1084 (1977).
- Fu, J., Zhang, H., Guo, F., Ma, L., Wu, J., Yue, M., Zheng, X., Qiu, Z., Li, L. Identification and characterization of abundant repetitive sequences in *Allium cepa*. *Sci. Rep.* **9**, 16756 (2019).

- Garrido-Ramos, M. A. Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* **146**, 153-170 (2015).
- Garrido-Ramos, M. A. Satellite DNA: an evolving topic. *Genes* **8**, 1-41 (2017).
- Hancock, J. M. Simple sequences and the expanding genome. *Bioessays* **18**, 421-425 (1996).
- Heslop-Harrison, J. S. P., Schwarzacher, T. Organisation of the plant genome in chromosomes. *Plant J.* **66**, 18–33 (2011).
- Hinegardner, R., Rosen, D. E. Cellular DNA content and evolution of Teleostean fishes. *Am. Nat.* **106**, 621-644 (1972).
- Jesus, C. M., Galetti Jr, P. M., Valentini, S. R., Moreira-filho, O. Molecular characterization and chromosomal localization of two families of satellite DNA in *Prochilodus Lineatus* (Pisces, Prochilodontidae), a species with B chromosomes. *Genetica* **118**, 25-32 (2003).
- Kit, S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *J. Mol. Biol.* **3**, 711-716 (1961).
- Koehler, M. R., Haaf, T., Guttenbach, M., Schartl, M., Schmid, M. Cytogenetics of the genus *Leporinus* (Pisces, Anostomidae). II. Molecular cytogenetics, organization and evolutionary conservation of a chromosome-specific satellite DNA from *Leporinus obtusidens*. *Chromosome Res.* **5**, 325-331 (1997).
- Lanfredi, M., Congiu, L., Garrido-Ramos, M. A., De La Herrán, R., Leis, M., Chicca, M., Rossi, R., Tagliavini, J., Rejón, C. R., Rejón, M. R., Fontana, F. Chromosomal location and evolution of a satellite DNA family in seven sturgeon species. *Chromosome Res.* **9**, 47-52 (2001).
- Lima, L. G., Svartman, M., Kuhn, G. C. S. Dissecting the satellite DNA landscape in three cactophilic *Drosophila* sequenced genomes. *G3 (Bethesda)* **7**, 2831-2843 (2017).
- López-Flores, I., Garrido-Ramos, M. A. The repetitive DNA content of eukaryotic genomes. In: Garrido-ramos, M. A. (Org.). Repetitive DNA. *Granada: Genome Dynamics, Karger, Basel* 1-28 (2012).
- Lovejoy, N. R., Lester, K., Crampton, W. G. R., Marques, F. P. L., Albert, J. S. Phylogeny, biogeography, and electric signal evolution of Neotropical knifefishes of the genus *Gymnotus* (Osteichthyes: Gymnotidae). *Mol. Phylogenet. Evol.* **54** 278-290 (2010).
- Louzada, S., Lopes, M., Ferreira, D., Adegas, F., Escudeiro, A., Gama-Carvalho, M., Chaves, R. Decoding the role of satellite DNA in genome architecture and plasticity—an evolutionary and clinical affair. *Genes* **11**, 72 (2020).
- Mago-Leccia, F. Electric fishes of the continental waters of America. *Caracas: Biblioteca De La Academia De Ciencias Fisicas, Matematicas Y Naturales* 206 (1994).

- Macas, J., Neumann, P., Navratilova, A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* **8**, 427 (2007).
- Macas, J., Kejnovsky, E., Neumann, P., Novák, P., Koblizkova, A., Vyskot, B. Next Generation Sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. *PLoS ONE* **6**, e27335 (2011).
- Martins, C., Ferreira, I. A., Oliveira, C., Foresti, F., Galetti Jr, P. M. A tandemly repetitive centromeric DNA sequence of the fish *Hoplias malabaricus* (Characiformes: Erythrinidae) is derived from 5S rDNA. *Genetica* **127**, 133-141 (2006).
- Mcmurray, C. T. Mechanisms of DNA expansion. *Chromosoma* **104** 2-13 (1995).
- Melo S., Utsunomia, R., Penitente, M., Sobrinho-scudeler, P. E., Porto-foresti, F., Oliveira, C., Foresti, F., Dergam, J. A. B Chromosome dynamics in *Prochilodus costatus* (Teleostei, Characiformes) and comparisons with supernumerary chromosome system in other *Prochilodus* species. *Comp. Cytogen.* **11**, 393-403 (2017).
- Mestriner, C. A., Galetti, P. M., Valentini, S. R., Ruiz, I. R. G., Abel, L. D. S., Moreira-Filho, O., Camacho, J. P. M. Structural and functional evidence that a B chromosome in the characid fish *Astyanax scabripinnis* is an isochromosome. *Heredity* **85**, 1-9 (2000).
- Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko T., Šatović, E., Plohl, M. Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Res.* **23**, 583 (2015).
- Milani, D., Cabral-de-Mello, D.C. Microsatellite organization in the grasshopper *Abracris flavolineata* (Orthoptera: Acrididae) revealed by fish mapping: remarkable spreading in the A and B chromosomes. *PLoS ONE* **9** (2014).
- Milhomem, S. S. R., Pieczarka, J. C., Crampton, W. G. R., Silva, D. S., Souza, A. C. P., Carvalho Jr., J. R., Nagamachi, C. Y. Chromosomal evidence for a cryptic species in the *Gymnotus carapo* species-complex (Gymnotiformes, Gymnotidae). *BMC Evol. Biol.* **9**, 75 (2008).
- Milhomem, S. S. R., Crampton, W. G. R., Pieczarka, J. C., Shetka, G. H., Silva, D. S., Nagamachi, C. Y. *Gymnotus capanema*, a new species of electric knife fish (Gymnotiformes, Gymnotidae) from eastern Amazonia, with comments on an unusual karyotype. *J. Fish Biol.* **80**, 812-815 (2012a).
- Milhomem, S. S. R., Crampton, W. G. R., Pieczarka, J. C., Silva, D. S., Cardoso, A. L., Silva, P. C., Oliveira, J. A., Nagamachi, C. Y. Chromosomal and electric signal diversity in three

- sympatric electric knifefish species (*Gymnotus*, Gymnotidae) from the central Amazon floodplain. *Rev. Fish Biol. Fish.* **22**, 485-497 (2012b).
- Milhomem, S. S. R., Scacchetti, P. C., Pieczarka, J. C., Ferguson-Smith, M. A., Pansonato-Alves, J. C., O'brien, P. C. M., Foresti, F., Nagamachi, C. Y. Are NORs always located on homeologous chromosomes? A fish investigation with rDNA and whole chromosome probes in *Gymnotus* fishes (Gymnotiformes). *PLoS ONE* **8**, e55608 (2013).
- Nei, M., Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**, 121 (2005).
- Novák, P., Neumann, P., Macas, J. Graph-based clustering and characterization of repetitive sequences in Next-Generation Sequencing data. *BMC Bioinformatics* **11**, 378 (2010).
- Novák, P., Neumann, P., Pech, J., Steinhais, J., Macas, J. Repeatexplorer: A Galaxy-based WEB server for genome-wide characterization of eukaryotic repetitive elements from Next-Generation sequence reads. *Bioinformatics* **29**, 792-793 (2013).
- Novák, P., Robledillo, L. A., Koblzková, A., Vrbová, I., Neumann, P., Macas, J. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **45**, 1 (2017).
- Oliveira, C., Foresti, F., Hilsdorf, A. W. S. Genetics of Neotropical fish: from chromosomes to populations. *Fish Physiol. Biochem.* **35**, 81-100 (2009).
- Pagan, H. J., Macas, J., Novak, P., Mcculloch, E. S., Stevens, R. D., Ray, D. A. Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. *Genome Biol. Evol.* **4**, 575-585 (2012).
- Palacios-Gimenez, O. M., Dias, G. B., Lima, L. G., Ramos, É., Martins, C., Cabral-de-Mello, D. C. High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the Neo-Y chromosome of the cricket *Eneoptera rurinamensis*. *Sci. Rep.* **7**, 6422 (2017).
- Parise-Maltempi, P. P., Da Silva, E. L., Rens, W., Dearden, F., O'brien, P. C., Trifonov, V., Ferguson-smith, M. A. Comparative analysis of sex chromosomes in *Leporinus* species (Teleostei, Characiformes) using chromosome painting. *BMC Genetics* **14** 60 (2013).
- Pinkel, D., Straume, T., Gray, J. W. Cytogenetic analysis using quantitative, high sensitivity, Fluorescence Hybridization. *Proc. Natl. Acad. Sci. U S A* **83**, 2934-2938 (1986).
- Plohl, M., Mestrovic, N., Mravinac, B. Satellite DNA Evolution. In: Garrido-ramos, M. A. (Org.). Repetitive DNA. *Granada: Genome Dynamics, Karger, Basel* 126-152 (2012).
- Reis, R. E., Albert, J. S., Dario, F. D., Mincarone, M. M., Petry, P., Rocha, L. A. Fish biodiversity and conservation in South America. *J. Fish Biol.* **90**, 1182 (2016).

- Robles, F., De La Herrán, R., Ludwig, A., Rejón, C. R., Rejón, M. R., Garrido-Ramos, M. A. Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* **338**, 133-142 (2004).
- Ruiz-Ruano, F. J., Cuadrado, Á., Montiel, E. E., Camacho, J. P. M., López-León, M. D. Next Generation Sequencing and FISH reveal uneven and nonrandom microsatellite distribution in two grasshopper genomes. *Chromosoma* **124**, 221-234 (2015).
- Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., Camacho, J. P. M. High-throughput analysis of the Satellitome illuminates satellite DNA evolution. *Sci. Rep.* **6** (2016a).
- Ruiz-Ruano, F. J., Cabrero, J., López-León, M. D., Camacho, J. P. M. Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. *Chromosoma* **124**, 487-500 (2016b).
- Ruiz-Ruano, F.J., Cabrero, J., López-León, M.D., Sánchez, A., Camacho, J. P. M. Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the migratory locust. *Chromosoma* **127**, 45 (2018).
- Santana, C. D., Crampton, W. G., Dillman, C. B., Frederico, R. G., Sabaj, M. H., Covain, R., Ready, J. S., Zuanon, J., Oliveira, R. R., Mendes-Junior, R. N., Bastos, D. A., Teixeira, T. F., Mol, J., Ohara, W. M., Castro, N. C., Peixoto, L. A., Nagamachi, C. Y., Sousa, L., Montag, L. F., Ribeiro, F., Waddell, J. C., Piorsky, N. M., Vari, R. P., Wosiacki, W. B. Unexpected species diversity in electric eels with a description of the strongest living bioelectricity generator. *Nature* (2019).
- Scacchetti, P. C., Pansonato-Alves, J. C., Utsunomia, R., Oliveira, C., Foresti, F. Karyotypic diversity in four species of the genus *Gymnotus* Linnaeus, 1758 (Teleostei, Gymnotiformes, Gymnotidae): physical mapping of ribosomal genes and telomeric sequences. *Comp. Cytogen.* **5**, 223-235 (2011).
- Scacchetti, P. C., Alves, J. C. P., Utsunomia, R., Claro, F. L., Almeida-Toledo, L. F., Oliveira, C., Foresti, F. Molecular characterization and physical mapping of two classes of 5S rDNA in the genomes of *Gymnotus sylvius* and *G. inaequilabiatus* (Gymnotiformes, Gymnotidae). *Cytogenet. Genome Res.* **136**, 131-137 (2012).
- Serrano-Freitas, E. A., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Utsunomia, R., Araya-Jaime, C., Oliveira, C., Camacho, J. P. M., Foresti, F. Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. *Mol. Genet. Genomics* **295**, 195–207 (2020).
- Silva, D. M. Z. A., Utsunomia, R., Ruiz-Ruano, F. J., Daniel, S. N., Porto-Foresti, F., Hashimoto, D. T., Oliveira, C., Camacho, J. P. M., Foresti, F. High-throughput analysis

- unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci. Rep.* **7**, 1 (2017).
- Silva, M., Matoso, D.A., Artoni, R. F., Feldberg, E. Karyotypic diversity and evolutionary trends in Neotropical electric fish of the genus *Gymnotus* (Gymnotiformes: Gymnotidae). *Zebrafish* **16**,308-320 (2019).
- Singer, M. F.Highly Repeated sequences in mammalian genomes. *Int. Rev. Cytol.* **76**, 67-112 (1982).
- Smit, A. F. A., Hubley, R., Green, P. RepeatMasker Open-3.0. (2010). Disponível Em: <<http://www.repeatmasker.org>>
- Steflova, P., Tokan, V., Vogel, I., Lexa, M., Macas, J., Novak, P., Hobza, R., Vyskot, B., Kejnovsky,E. Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY1Y2) in the dioecious plant *Rumex acetosa*. *Genome Biol. Evol.* **5**, 769-782 (2013).
- Steinemann, S., Steinemann, M. Y Chromosomes: born to be destroyed. *Bioessays* **27**, 1076-1083 (2005).
- Thomas Jr, C. A. The genetic organization of chromosomes. *Annu. Rev. Genet.* **5**, 237-256 (1971).
- Utsunomia, R., Pansonato-Alves, J. P., Paiva, R., Costa-Silva, G., Oliveira, C., Bertollo, L., Foresti, F. Genetic differentiation among distinct karyomorphs of the wolf fish *Hoplias malabaricus* species complex (Characiformes, Erythrinidae) and report of unusual hybridization with natural triploidy. *J. Fish Biol.* **85** (2014).
- Utsunomia, R., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Araya-Jaime, C., Pansonato-Alves, J. C., Scacchetti, P. C., Hashimoto, D. T., Oliveira, C., Trifonov, V. A., Porto-Foresti, F., Camacho, J. P. M., Foresti, F. Uncovering The Ancestry Of B Chromosomes In *Moenkhausia sanctaefilomenae* (Teleostei, Characidae). *PLoS ONE* **1**, 1-20 (2016).
- Utsunomia, R., Ruiz-Ruano, F. J., Silva, D. M. Z. A., Serrano, E. A., Rosa, I. F., Sobrinho-Scudeler, P. E., Hashimoto, D. T., Oliveira, C., Camacho, J. P. M., Foresti, F. A glimpse into the Satellite DNA library in Characidae fish (Teleostei, Characiformes). *Front. Genet.* **8**, 103 (2017).
- Utsunomia, R., Melo, S., Scacchetti, P. C., Oliveira, C., Machado, M. A, Pieczarka, J. C., Nagamachi C. Y., Foresti, F. Particular Chromosomal Distribution of Microsatellites in Five Species of the Genus *Gymnotus* (Teleostei, Gymnotiformes). *Zebrafish* **15**, 398-403 (2018).

- Utsunomia, R., Silva, D. M. Z. A., Ruiz-Ruano, F.J., Goes, C. A. G., Melo, S., Ramos, L. F. P., Oliveira, C., Porto-Foresti, F., Foresti, F., Hashimoto, D. T. Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* **9**, 5856 (2019).
- Tagliacollo, V. A., Bernt, M. J., Craig, J. M., Oliveira, C., Albert, J. S. Model-based total evidence phylogeny of Neotropical electric knifefishes (Teleostei, Gymnotiformes). *Mol. Phylogenet. Evol.* **95**, 20-33 (2016).
- Van der Sleen. P., Albert, J. S. Field guide to the fishes of the Amazon, Orinoco, and Guianas. *University of Princeton Press* (2017).
- Vicari, M. R., Nogaroto, V., Noleto R.B., Cestari, M. M., Cioffi, M. B., Almeida, M. C., Moreira-Filho, O., Bertollo, L. A. C., Artoni, R. F. Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. *J. Fish Biol.* **76**, 1094-1116 (2010).

7 MATERIAL SUPLEMENTAR

7.1 Material Suplementar – Capítulo 1

Os satelitomas em .fasta de *Gymnotus cuia* e *Gymnotus sylvius* podem ser acessados no seguinte endereço eletrônico:

< https://drive.google.com/drive/folders/1O4bjS-yKnZwgjrisklAn_3IW8WxI5RE?usp=sharing >

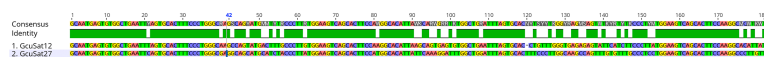
Tabela Suplementar 1.1. Famílias de DNA satélite de *Gymnotus cuia* com indicação de Superfamílias (SF), satélites GcuSats-tamanho das Unidades de Repetição (UR), porcentagem A+T, abundância, porcentagem de divergência (Kimura 2-parâmetros) e número de variantes.

SF	GcuSat	A + T (%)	Abundância	Divergência (%)	Variantes
	GcuSat01-118	71,2	0,134575	2,19	2
	GcuSat02-611	63,8	0,133884	5,45	1
	GcuSat03-537	54,2	0,058035	1,99	1
	GcuSat04-221	61,1	0,055741	4,07	1
	GcuSat05-368	56,5	0,051165	3,05	2
	GcuSat06-592	55,9	0,045575	5,16	1
	GcuSat07-283	59,7	0,044621	3,88	1
	GcuSat08-262	55,3	0,031879	4,85	1
	GcuSat09-177	49,7	0,025430	5,1	1
	GcuSat10-1588	46,7	0,023406	3,77	1
	GcuSat11-1006	50,2	0,022617	3,74	1
A'	GcuSat12-181	53,0	0,022270	8,13	1
	GcuSat13-1221	57,6	0,020993	12,38	1
	GcuSat14-1248	54,9	0,020118	10,88	1
	GcuSat15-703	60,2	0,020070	1,28	1
	GcuSat16-1015	56,2	0,017514	1,71	1
	GcuSat17-118	74,6	0,013561	7,72	1
	GcuSat18-216	62,0	0,011085	2,63	1
	GcuSat19-64	71,9	0,010661	8,91	1
	GcuSat20-1205	59,4	0,010466	7,34	1
	GcuSat21-1224	63,7	0,009458	0,74	1
	GcuSat22-1365	59,8	0,007761	3,7	1
	GcuSat23-303	54,1	0,007596	11,08	1
	GcuSat24-34	38,2	0,007140	4,46	1
	GcuSat25-263	59,3	0,006587	8,73	1
	GcuSat26-685	58,7	0,006124	1,76	1

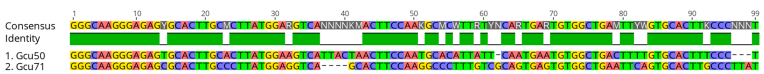
A'	GcuSat27-182	50,0	0,005599	13,6	1
	GcuSat28-45	53,3	0,005542	4,47	1
	GcuSat29-1052	63,1	0,005184	3,77	1
	GcuSat30-107	51,4	0,004988	4,41	2
	GcuSat31-2109	65,6	0,004939	4,47	1
	GcuSat32-1431	55,8	0,004444	14,49	1
	GcuSat33-1037	55,8	0,004418	2,66	1
	GcuSat34-1110	55,6	0,004300	2,63	1
	GcuSat35-1992	55,6	0,004266	0,91	1
	GcuSat36-42	59,5	0,003991	4,08	1
	GcuSat37-2079	51,3	0,003870	3,11	1
	GcuSat38-147	38,8	0,003658	4,42	1
	GcuSat39-925	61,4	0,003522	1,1	1
	GcuSat40-1511	63,7	0,003419	2,59	1
	GcuSat41-1020	55,7	0,003411	6,63	1
	GcuSat42-107	54,2	0,003411	3,18	1
	GcuSat43-51	49,0	0,003337	8,17	1
	GcuSat44-186	56,5	0,003077	10,02	1
	GcuSat45-172	72,7	0,003008	2,37	1
	GcuSat46-1154	54,8	0,002951	2,34	1
	GcuSat47-338	56,5	0,002766	3,36	1
	GcuSat48-177	61,6	0,002610	9,51	1
	GcuSat49-831	58,7	0,002608	1,38	1
B'	GcuSat50-95	56,8	0,002561	3,66	2
	GcuSat51-257	54,1	0,002490	8,83	1
	GcuSat52-768	63,8	0,002356	1,4	1
	GcuSat53-892	56,8	0,002347	4,94	1
	GcuSat54-1640	59,2	0,002057	4,37	1
	GcuSat55-847	62,7	0,001939	1,61	1
	GcuSat56-42	59,5	0,001748	7,1	1
	GcuSat57-825	59,5	0,001702	3,19	1
	GcuSat58-190	67,4	0,001694	2,17	1
	GcuSat59-49	40,8	0,001618	4,78	1
	GcuSat60-756	57,5	0,001465	1,65	1
	GcuSat61-581	55,9	0,001387	3,2	1
	GcuSat62-84	56,0	0,001363	12,99	1
	GcuSat63-726	55,4	0,001328	8,64	1
	GcuSat64-2079	51,3	0,001279	7	1
C'	GcuSat65-239	60,3	0,001219	2,91	1
	GcuSat66-1456	66,9	0,001192	8,12	1
	GcuSat67-633	57,0	0,001165	12,09	1
	GcuSat68-1538	61,2	0,001159	5,51	1
	GcuSat69-454	50,0	0,001120	1,79	1

	GcuSat70-243	60,1	0,001028	3,68	1
B'	GcuSat71-95	44,2	0,000957	17,54	1
	GcuSat72-1820	65,2	0,000930	0,98	1
	GcuSat73-186	68,8	0,000813	2,51	1
	GcuSat74-1236	64,2	0,000804	1,22	1
	GcuSat75-411	57,7	0,000791	10,15	1
D'	GcuSat76-764	66,0	0,000758	6,79	1
	GcuSat77-345	58,0	0,000757	6,84	1
	GcuSat78-767	66,1	0,000704	2,74	1
	GcuSat79-971	60,5	0,000678	3,63	1
	GcuSat80-174	54,0	0,000644	5,34	1
C'	GcuSat81-239	64,4	0,000619	2,08	1
	GcuSat82-493	57,8	0,000594	3,02	1
	GcuSat83-105	55,2	0,000483	5,34	1
	GcuSat84-37	54,1	0,000455	4,61	1
	GcuSat85-40	87,5	0,000448	2,17	1
	GcuSat86-1194	54,8	0,000431	12,84	1
	GcuSat87-214	72,0	0,000372	7,53	1
	GcuSat88-228	66,7	0,000350	14,1	1
	GcuSat89-973	60,4	0,000317	5,42	1
	GcuSat90-493	58,0	0,000235	6,69	1
D'	GcuSat91-764	65,6	0,000190	0,42	1
	GcuSat92-45	53,3	0,000082	6,59	1

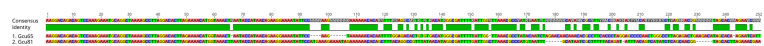
SFA



SFB



SFC



SFD



Figura Suplementar 1.1. Alinhamento das seqüências satélites de *G. cuia* caracterizadas como SuperFamílias (SF).

Tabela Suplementar 1.2. Famílias de DNA satélite de *Gymnotus sylvius* com indicação de Superfamílias (SF), satélites GsySats-tamanho das Unidades de Repetição (UR), porcentagem A+T, abundância, porcentagem de divergência (Kimura 2-parâmetros) e número de variantes.

SF	GsySat	A+T (%)	Abundância	Divergência (%)	Variantes
	GsySat01-216	59,3	0,215465	2,5	2
	GsySat02-602	59,5	0,028718	2,65	1
	GsySat03-139	63,3	0,022923	8,8	1
	GsySat04-1186	55,7	0,021832	10,34	1
	GsySat05-1248	55,3	0,017440	12,8	1
	GsySat06-1270	47,2	0,017212	6,37	1
	GsySat07-181	51,4	0,016511	10,42	2
	GsySat08-176	72,7	0,015282	3,27	1
	GsySat09-1126	55,7	0,014969	15,23	1
	GsySat10-1038	36,8	0,012917	11,16	1
	GsySat11-235	60,0	0,010859	13,8	1
	GsySat12-83	43,4	0,010115	5,26	1
	GsySat13-706	54,5	0,009514	5,33	1
	GsySat14-1225	63,7	0,009484	1,28	1
	GsySat15-34	38,2	0,008368	4,71	1
	GsySat16-810	57,4	0,008173	1,88	1
	GsySat17-1392	56,4	0,006958	2,04	1
A	GsySat18-107	52,3	0,006737	4,45	1
	GsySat19-1204	59,1	0,006640	6,04	1
	GsySat20-45	55,6	0,006597	4,94	1
A	GsySat21-99	52,5	0,006183	6,44	1
A	GsySat22-95	55,8	0,005836	6,92	1
	GsySat23-512	57,2	0,005770	7,43	1
	GsySat24-1384	55,3	0,005019	3,47	1
	GsySat25-1083	59,3	0,004894	2,39	1
	GsySat26-42	33,3	0,004272	15,7	1
	GsySat27-1400	59,4	0,004131	4,95	1
	GsySat28-1568	63,7	0,004053	4,29	1
	GsySat29-1865	58,7	0,003690	1,45	1
	GsySat30-257	55,6	0,003105	3,9	2
	GsySat31-809	57,6	0,003090	2,35	1
	GsySat32-49	40,8	0,003035	5,44	1
	GsySat33-1153	54,1	0,002970	3,43	1
	GsySat34-1036	56,3	0,002944	3,17	1
	GsySat35-42	59,5	0,002852	5,21	1
	GsySat36-1954	55,3	0,002691	0,84	1
	GsySat37-36	27,8	0,002408	13,55	1

	GsySat38-855	55,9	0,002260	6,32	1
	GsySat39-798	62,8	0,002129	1,65	1
	GsySat40-163	62,0	0,002075	4,68	1
	GsySat41-72	43,1	0,002017	4,66	1
	GsySat42-30	60,0	0,001921	6,16	1
	GsySat43-180	53,9	0,001814	3,76	1
	GsySat44-38	50,0	0,001771	12,95	1
	GsySat45-255	56,9	0,001721	10,97	1
	GsySat46-346	55,2	0,001719	6,12	1
	GsySat47-803	58,7	0,001645	2,7	1
	GsySat48-84	56,0	0,001554	13,32	1
	GsySat49-756	56,5	0,001552	2,63	1
	GsySat50-76	56,6	0,001523	3,63	1
	GsySat51-1117	49,8	0,001520	2,31	1
	GsySat52-54	46,3	0,001514	4,51	1
	GsySat53-66	56,1	0,001496	4,71	1
	GsySat54-220	46,8	0,001424	4,31	1
	GsySat55-42	64,3	0,001423	3,61	1
	GsySat56-245	59,6	0,001310	3,34	1
	GsySat57-166	67,5	0,001296	1,75	1
	GsySat58-848	62,4	0,001294	2,2	1
	GsySat59-350	62,3	0,001205	8,56	1
	GsySat60-486	51,9	0,001116	9,83	1
	GsySat61-174	60,3	0,001077	7,87	1
	GsySat62-1034	59,6	0,001055	1,24	1
	GsySat63-183	57,4	0,000949	4,27	2
B	GsySat64-165	56,4	0,000909	7,72	1
	GsySat65-110	66,4	0,000899	9,85	1
	GsySat66-64	71,9	0,000767	10,31	1
	GsySat67-43	60,5	0,000665	4,02	1
	GsySat68-99	37,4	0,000573	2,92	1
B	GsySat69-165	56,4	0,000441	6,87	1

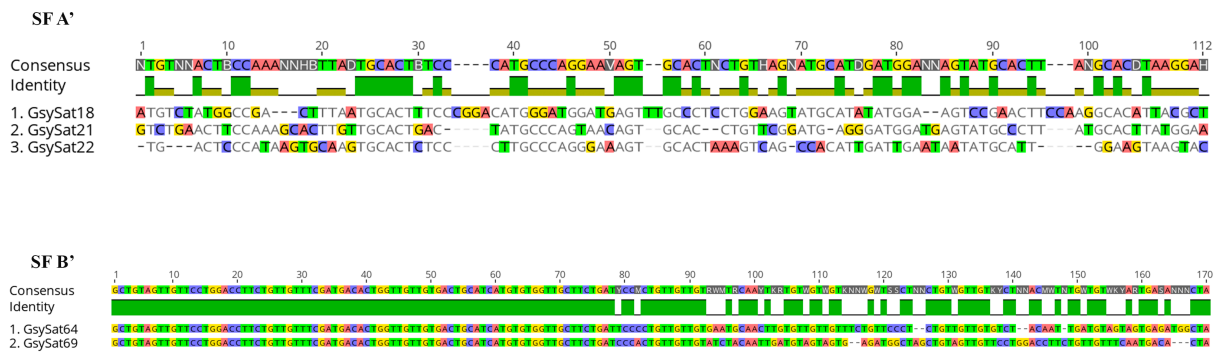


Figura Suplementar 1.2. Alinhamento das sequências satélites de *G. sylvius* caracterizadas como SuperFamílias (SF).

6.2 Material Suplementar – Capítulo 2

Os satelitomas em .fasta de *G. javari* (complexo *G. coatesi*), *G. inaequilabiatus* (complexo *G. tigre*) e *G. cylindricus* (complexo *G. cylindricus*) podem ser acessados no seguinte endereço eletrônico:

< https://drive.google.com/drive/folders/1O4bjS-yKnZwgjrisklAn_3lW8WxI5RE?usp=sharing >

Tabela Suplementar 2.1. Famílias de DNA satélite de *Gymnotus javari* com satélites GjaSats-tamanho das Unidades de Repetição (UR), porcentagem A+T, abundância, porcentagem de divergência (Kimura 2-parâmetros) e número de variantes.

GjaSat	A + T (%)	Abundância	Divergência (%)	Variantes
GjaSat01-206	65,0	0,178966	3,8	1
GjaSat02-186	65,1	0,074238	2,45	2
GjaSat03-93	54,8	0,067607	9,54	4
GjaSat04-65	53,8	0,050719	4,14	2
GjaSat05-251	58,2	0,047732	2,67	1
GjaSat06-1476	55,8	0,035351	14,8	1
GjaSat07-1026	61,2	0,029512	3,58	1
GjaSat08-950	56,6	0,028233	18,2	1
GjaSat09-2796	56,7	0,027258	9,14	1
GjaSat10-270	54,8	0,020899	5,2	1

GjaSat11-1639	59,7	0,016116	4,29	1
GjaSat12-486	56,0	0,015521	10,67	1
GjaSat13-1397	55,8	0,013858	9,77	1
GjaSat14-876	59,9	0,010743	3,4	1
GjaSat15-1009	57,9	0,009409	2,08	1
GjaSat16-98	54,1	0,008784	4,34	1
GjaSat17-1238	54,3	0,008373	1,27	1
GjaSat18-1223	63,7	0,008346	1,76	1
GjaSat19-49	42,9	0,007583	7,41	2
GjaSat20-685	53,3	0,006717	14,69	1
GjaSat21-1505	55,5	0,006688	5,93	1
GjaSat22-38	52,6	0,006494	6,21	1
GjaSat23-74	68,9	0,005984	2,75	1
GjaSat24-1436	55,6	0,005972	1,4	1
GjaSat25-1045	58,8	0,005861	3,49	1
GjaSat26-89	44,9	0,005723	7,03	2
GjaSat27-1161	52,3	0,005420	5,08	1
GjaSat28-1959	59,5	0,005417	3,37	1
GjaSat29-2009	51,3	0,005088	1,52	1
GjaSat30-789	48,0	0,005071	1,88	1
GjaSat31-56	39,3	0,004992	9,75	1
GjaSat32-84	56,0	0,004829	14,33	1
GjaSat33-76	56,6	0,004296	4,59	1
GjaSat34-402	54,0	0,004235	5,38	1
GjaSat35-837	62,6	0,004145	3,49	1
GjaSat36-33	57,6	0,004111	5,24	1
GjaSat37-92	58,7	0,003940	2,41	1
GjaSat38-345	55,7	0,003796	7,02	1
GjaSat39-1149	54,0	0,003368	4,52	1
GjaSat40-35	62,9	0,003202	4,76	1
GjaSat41-759	61,0	0,002974	4,69	1
GjaSat42-174	60,3	0,002749	5,72	1
GjaSat43-458	60,3	0,002611	4,33	1
GjaSat44-887	62,7	0,002608	4,11	1
GjaSat45-306	51,3	0,002508	10,04	1
GjaSat46-776	64,4	0,002458	0,99	1
GjaSat47-258	56,2	0,002126	7,95	1
GjaSat48-37	45,9	0,002001	12,63	1
GjaSat49-186	69,9	0,001895	5,72	1
GjaSat50-135	39,3	0,001820	11,27	2
GjaSat51-51	47,1	0,001814	8,61	1
GjaSat52-32	56,3	0,001770	8,83	1
GjaSat53-46	50,0	0,001737	6,38	1

GjaSat54-96	66,7	0,001698	2,96	1
GjaSat55-42	59,5	0,001532	7,16	1
GjaSat56-200	60,0	0,001488	7,76	1
GjaSat57-745	57,4	0,001157	1,85	1
GjaSat58-41	53,7	0,001125	7,31	1
GjaSat59-64	57,8	0,001075	4,91	1
GjaSat60-661	59,9	0,001073	3,57	1
GjaSat61-69	78,3	0,001019	1,46	1
GjaSat62-42	64,3	0,000933	2,58	1
GjaSat63-41	56,1	0,000877	6,92	1
GjaSat64-79	53,2	0,000191	4,25	1

Tabela Suplementar 2.2. Famílias de DNA satélite de *Gymnotus inaequilabiatus* com satélites GinSats-tamanho das Unidades de Repetição (UR), porcentagem A+T, abundância, porcentagem de divergência (Kimura 2-parâmetros) e número de variantes.

GinSat	A + T (%)	Abundância	Divergência (%)	Variantes
GinSat01-216	62,5	0,106008	7,06	1
GinSat02-118	71,2	0,102165	6,41	1
GinSat03-603	55,6	0,036598	7,23	1
GinSat04-931	59,1	0,029115	1,4	1
GinSat05-453	68,4	0,026445	7,03	1
GinSat06-177	49,2	0,025491	7,17	1
GinSat07-1248	54,7	0,025379	10,99	1
GinSat08-702	47,2	0,019139	8,57	1
GinSat09-91	51,6	0,017423	12,98	2
GinSat10-168	64,9	0,015407	12,27	1
GinSat11-568	50,9	0,014593	17,98	1
GinSat12-1205	59,5	0,008639	7,8	1
GinSat13-83	45,8	0,008489	5,22	5
GinSat14-107	51,4	0,007976	8,25	2
GinSat15-34	38,2	0,007592	4,93	2
GinSat16-2079	51,3	0,007262	2,65	1
GinSat17-1403	59,2	0,006661	3,94	1
GinSat18-396	58,1	0,006600	1,31	1
GinSat19-1110	55,5	0,006582	3,25	1
GinSat20-1224	63,9	0,006206	0,69	1
GinSat21-45	55,6	0,005719	4,81	1
GinSat22-194	57,7	0,004987	10,76	2
GinSat23-1027	56,9	0,003942	2,57	1
GinSat24-31	54,8	0,003641	4,8	1

GinSat25-42	57,1	0,003583	6,74	2
GinSat26-752	58,0	0,003542	1,41	1
GinSat27-700	55,1	0,003265	5,14	1
GinSat28-95	56,8	0,003164	4,58	1
GinSat29-257	56,0	0,002884	8,32	1
GinSat30-581	55,8	0,002460	2,77	1
GinSat31-847	62,8	0,002412	1,49	1
GinSat32-1152	55,1	0,002349	4,1	1
GinSat33-49	42,9	0,002316	5,78	1
GinSat34-171	72,5	0,002073	3,43	1
GinSat35-454	50,0	0,001956	1,54	1
GinSat36-54	38,9	0,001943	7,13	2
GinSat37-936	61,4	0,001904	1,32	1
GinSat38-798	63,0	0,001716	1,44	1
GinSat39-42	59,5	0,001380	4,18	1
GinSat40-708	54,2	0,001244	1,95	1
GinSat41-76	56,6	0,001188	4,01	1
GinSat42-36	52,8	0,001031	4,9	1
GinSat43-38	52,6	0,000893	8,23	1
GinSat44-186	69,9	0,000692	2,37	1
GinSat45-175	53,7	0,000680	4,53	1
GinSat46-85	52,9	0,000471	4,1	1
GinSat47-119	67,2	0,000469	1,08	2
GinSat48-64	59,4	0,000404	3,09	1
GinSat49-125	68,0	0,000250	2,92	1

Tabela Suplementar 2.3. Famílias de DNA satélite de *Gymnotus cylindricus* com satélites GcySats-tamanho das Unidades de Repetição (UR), porcentagem A+T, abundância, porcentagem de divergência (Kimura 2-parâmetros) e número de variantes.

GcySat	A + T (%)	Abundância	Divergência (%)	Variantes
GcySat01-175	59,4	0,130613	6,5	1
GcySat02-51	58,8	0,068412	1,9	1
GcySat03-35	62,9	0,054661	7,1	3
GcySat04-595	60,3	0,042802	1,87	1
GcySat05-715	44,5	0,024399	3,78	1
GcySat06-104	56,7	0,019423	6,98	1
GcySat07-1224	64,1	0,014673	2,85	1
GcySat08-1361	55,7	0,013888	6,33	1
GcySat09-1049	57,6	0,011065	10,97	1
GcySat10-1245	56,5	0,010453	6,06	1
GcySat11-1094	60,1	0,009767	4,21	1
GcySat12-1210	59,8	0,009284	6,95	1

GcySat13-32	53,1	0,008481	3,77	3
GcySat14-1801	53,6	0,008347	7,49	1
GcySat15-34	38,2	0,007921	5,48	1
GcySat16-834	58,9	0,007279	13,94	1
GcySat17-47	53,2	0,006734	4,78	2
GcySat18-282	53,5	0,005923	6,62	1
GcySat19-1378	56,7	0,005895	7,71	1
GcySat20-1033	58,4	0,004825	4,89	1
GcySat21-258	55,4	0,004444	5,24	2
GcySat22-33	48,5	0,004344	5,48	1
GcySat23-1131	48,7	0,003658	1,69	1
GcySat24-1151	53,5	0,003632	2,26	1
GcySat25-1126	56,9	0,003623	2,17	1
GcySat26-42	59,5	0,003370	4,62	1
GcySat27-1146	59,2	0,003098	3,68	1
GcySat28-82	50,0	0,003039	3,31	1
GcySat29-234	61,1	0,002942	10,54	1
GcySat30-852	62,6	0,002861	2,66	1
GcySat31-169	58,6	0,002730	4,51	1
GcySat32-490	64,1	0,002640	3,23	1
GcySat33-54	38,9	0,002533	7,38	1
GcySat34-336	55,7	0,002278	8,16	1
GcySat35-39	59,0	0,002253	6,3	2
GcySat36-781	64,0	0,002206	1,31	1
GcySat37-76	56,6	0,002032	3,79	1
GcySat38-171	71,9	0,001906	3,05	2
GcySat39-599	57,9	0,001904	2,62	1
GcySat40-527	58,8	0,001852	3,3	1
GcySat41-42	59,5	0,001744	4,1	1
GcySat42-64	71,9	0,001613	8,4	2
GcySat43-34	55,9	0,001598	4,62	1
GcySat44-337	62,6	0,001593	8,61	1
GcySat45-55	43,6	0,001442	5,07	1
GcySat46-345	57,4	0,001401	5,62	1
GcySat47-38	63,2	0,001371	2,86	1
GcySat48-36	50,0	0,001103	4,73	1
GcySat49-29	48,3	0,000972	7,57	3
GcySat50-35	57,1	0,000949	6,68	2
GcySat51-39	48,7	0,000910	7,99	1
GcySat52-63	66,7	0,000900	3,28	2
GcySat53-42	47,6	0,000831	3,82	1
GcySat54-43	46,5	0,000697	1,46	3

8 ANEXOS

8.1 Anexo I – Certificado do Protocolo de ética CEUA 974/2017 (IBB/UNESP)



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



Certificado

Certificamos que o projeto intitulado “ANÁLISES CROMOSSÔMICAS E GENÔMICAS APLICADAS AO ESTUDO EVOLUTIVO E ESTRUTURAL DOS SATELITOMAS EM ESPÉCIES DO GÊNERO *Gymnotus* (TELEOSTEI, GYMNOTIFORMES)”, Protocolo nº 974-CEUA, sob a responsabilidade de **Fausto Foresti**, que envolve a produção, manutenção e/ou utilização de animais pertencentes ao filo Chordata, subfilo Vertebrata (exceto o homem), para fins de pesquisa científica (ou ensino) – encontra-se de acordo com os preceitos da Lei nº 11.794, de 9 de outubro de 2008, do Decreto nº 6.899, de 15 de julho de 2009, e com as normas editadas pelo Conselho Nacional de Controle da Experimentação Animal (CONCEA), e foi aprovado *ad referendum* da **COMISSÃO DE ÉTICA NO USO DE ANIMAIS** (CEUA), nesta data.

Finalidade:	<input type="checkbox"/> Ensino	<input checked="" type="checkbox"/> Pesquisa Científica
Vigência do Projeto:	Início: 22/5/2017	Término: 01/3/2020
Espécie/linhagem:	<i>Peixe Gymnotus inaequilabiatu</i> s e <i>Gymnotus sylvius</i>	
Nº de animais:	60	
Peso:	100g	Idade: sem previsão
Sexo:	Macho e fêmea	
Origem	Cachoeira Véu da Noiva – Município de Botucatu/SP	

Botucatu, 17 de maio de 2017.

Prof. Dr. Bruno Cesar Schimming
Presidente da CEUA

Instituto de Biociências – Diretoria Técnica Acadêmica
Distrito de Rubião Júnior/s/n CEP 18618-970 Botucatu SP Brasil
Tel 14 3880 0851 fax 14 3880 3744 e-mail: secda@ibb.unesp.br



8.2 Anexo II - RepeatExplorer utilizado x RepeatExplorer *TAREAN*

Os elementos repetitivos constituem uma proporção genômica considerável nas células animais e vegetais, podendo evidenciar padrões evolutivos que poderiam ser utilizados para a compreensão evolutiva de diversos grupos. A plataforma de filtragem de elementos repetitivos RepeatExplorer, proposta em 2013 por Novák et al. vem sendo amplamente utilizada para o isolamento de sequências repetitivas, especialmente de regiões de DNA satélite. Nessa ferramenta de filtragem, o isolamento das sequências se dá pela identificação dos gráficos com padrão circular e a extração manual dos *motifs* (unidades de repetição). Em 2017, Novák e colaboradores descreveram a plataforma *TAREAN* (*tandem repeat analyzer*), com a proposta de identificar prováveis sequências satélites, pontuar (*scoring*) os gráficos circulares através de algoritmos computacionais e já disponibilizar um arquivo de saída com as sequências consenso isoladas automaticamente (Figura Anexo II.1).

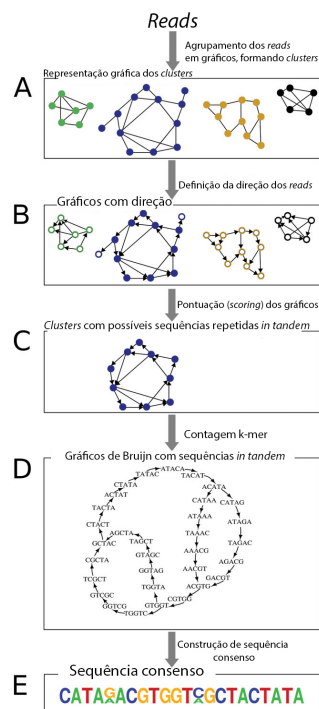


Figura Anexo II.1. O funcionamento da plataforma *TAREAN* consiste na análise dos mais variados tipos de gráficos circulares (A), estabelecendo uma direção pros *reads* e pontuando os gráficos (B). Os *shapes* circulares com maior pontuação são alocados em forma de Bruijn (gráficos com sequências direcionais) (D) e a sequência consenso desse gráfico é disponibilizada (E). (Modificado de Novák et al. 2017).

A plataforma *TAREAN* chegou a ser considerada no desenvolvimento da presente tese. Foram realizadas duas iterações na espécie *G. sylvius*, porém as sequências extraídas apresentaram monômeros de grande extensão, com média acima de 500 nt, e nenhum *motif* extraído apresentou menos de 165 nt. Nessas duas iterações, o total de 24 sequências foi isolado (Tabela Anexo II.1), o que foi considerado insuficiente e pouco diverso, visto que os satélites de monômeros menores não foram identificados.

Adicionalmente, foi possível observar que parte dos satélites mais abundantes dentro do genoma de *G. sylvius* (GsySat02, GsySat04, GsySat05, GsySat06) advieram dessas iterações iniciais, sugerindo que a plataforma *TAREAN* é eficaz em selecionar satélites mais abundantes presentes no genoma, todavia não é uma abordagem que representaria todos os potenciais satélites, como demonstrado na figura Anexo II.2. Visto que um dos objetivos principais da presente tese é construir e anotar a biblioteca com todos os possíveis satélites, sendo mais ou menos abundantes no genoma das espécies, optamos pela utilização da plataforma descrita por Novák et. al de 2013, com a verificação e seleção manuais das sequências satélites identificadas.

Tabela Anexo II.1. Satélites extraídos de duas iterações realizadas através da plataforma *TAREAN* (Novák et al. 2017), totalizando 24 sequências em duas iterações.

satDNA - <i>G. sylvius</i>	Run1	Run2
Menor monômero (nt)	180	166
Maior monômero (nt)	1400	1204
Média (nt)	934	537
A + T (%)	56,8	58,9
Total de satDNAs	19	05

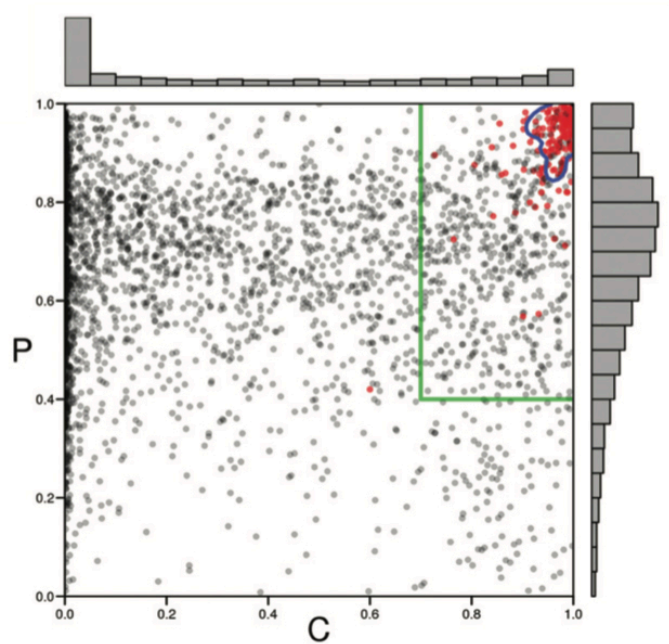


Figura Anexo II.2. Gráfico representando o modelo de predição de melhores satélites da plataforma *TAREAN*. A seleção de satélites de alta confiança se baseia modelo de análise discriminante, que é a correlação entre *C* (índice de conexão entre *reads*) e *P* (índice de similaridade de *reads* conectados). Os pontos cinzas representam todos os clusters de uma iteração e os pontos vermelhos marcam *clusters* que foram anotados manualmente como repetições de satélite. Os satélites de maior confiança que seriam disponibilizados pelo *TAREAN* estão dentro do limite da linha azul. A linha verde delimita e indica os *clusters* que possuem maior confiabilidade. (Modificado de Novák et al. 2017).

9 APÊNDICE

9.1 Produção Acadêmica e Científica no Doutorado (2016-2020)

Artigos Publicados

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

UTSUNOMIA, R.; **MELO, S.**; SCACCHETTI, P. C.; OLIVEIRA, C.; MACHADO, M. A.; PIECZARKA, J. C.; NAGAMACHI, C. Y.; FORESTI, F. Particular chromosomal distribution of microsatellites in five species of the genus *Gymnotus* (Teleostei, Gymnotiformes). *Zebrafish*, v.00, p.00, 2018.

CARVALHO*, M. L.; COSTA SILVA*, G. J.; **MELO***, S.; ASHIKAGA, F. Y.; SHIMABUKURO-DIAS, C. K.; SCACCHETTI, P. C.; DEVIDÉ, R.; FORESTI, F.; OLIVEIRA, C. The non-monotypic status of the Neotropical fish genus *Hemiodontichthys* (Siluriformes, Loricariidae) evidenced by genetic approaches. *Mitochondrial DNA Part A*, v.30, p.1 - 7, 2018. *These authors contributed equally for this work.

MELO, S.; UTSUNOMIA, R.; PENITENTE, M.; SOBRINHO-SCUDELER, P. E.; PORTO-FORESTI, F.; OLIVEIRA, C.; FORESTI, F.; DERGAM, J. A. B chromosome dynamics in *Prochilodus costatus* (Teleostei, Characiformes) and comparisons with supernumerary chromosome system in other *Prochilodus* species. *Comparative Cytogenetics*, v.11, p.393 - 403, 2017.

Informações Relevantes

- Resumos publicados em Anais de eventos: 16.

- *International Visiting Graduate Student* na *University of Toronto*: setembro/2018- fevereiro/2019. Supervisor: Nathan Lovejoy.
- Organização de eventos, exposições e feiras: **Workshop de Genética** (2019); **Pint of Science Brasil** – regional de Botucatu (2017, 2018, 2019); **Curso teórico-prático em Evolução Molecular** (2016).
- Participação no Projeto de Extensão Cascudinho: conservação da fauna aquática em riachos de Botucatu (2019).
- 17 participações em eventos.