

INSTITUTO DE BIOCÊNCIAS
Departamento de Biologia Estrutural e Funcional

**ESTUDOS CITOGENÉTICOS E GENÔMICOS EM *Spintherobolus papilliferus*
(Characiformes: Spintherobolidae): UMA ABORDAGEM INTEGRADA**

Letícia Batista Soares

Botucatu/SP

2025

Universidade Estadual Paulista “Júlio De Mesquita Filho”

Instituto de Biociências de Botucatu

**Estudos citogenéticos e genômicos em *Spintherobolus papilliferus* (Characiformes:
Spintherobolidae): uma abordagem integrada**

Aluno: Letícia Batista Soares

Orientador: Dr. Claudio de Oliveira

Defesa apresentada ao Instituto de Biociências,
Campus de Botucatu, UNESP, para obtenção do
título de Mestre pelo Programa de Pós-Graduação
em Ciências Biológicas (Genética).

Botucatu/SP

2025

S676e	Soares, Leticia Batista Estudos citogenéticos e genômicos em <i>Spintherobolus papilliferus</i> (Characiformes: Spintherobolidae): uma abordagem integrada / Leticia Batista Soares. -- Botucatu, 2025 43 f. : fotos Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, Botucatu Orientador: Cláudio de Oliveira 1. Ciências biológicas. 2. Peixes de água doce. 3. Citogenética animal. 4. Bioinformática. I. Título.
-------	--

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: ESTUDOS CITOGENÉTICOS E GENÔMICOS EM *Spintherobolus papilliferus*:
UMA ABORDAGEM INTEGRADA

AUTORA: LETICIA BATISTA SOARES

ORIENTADOR: CLAUDIO DE OLIVEIRA

Aprovada como parte das exigências para obtenção do Título de Mestra em Ciências Biológicas (Genética), pela Comissão Examinadora:

Prof. Dr. CLAUDIO DE OLIVEIRA (Participação Presencial)
Departamento de Biologia Funcional e Estrutural / Instituto de Biociências da Unesp de Botucatu

Prof. Dr. BRUNO OLIVEIRA DA SILVA DURAN (Participação Virtual)
Departamento de Histologia, Embriologia e Biologia Celular; Instituto de Ciências Biológicas (ICB) /
Universidade Federal de Goiás (UFG)

Prof. Dr. RICARDO UTSUNOMIA (Participação Presencial)
Departamento de Ciências Biológicas / Faculdade de Ciências - Unesp/Campus de Bauru

Botucatu, 26 de fevereiro de 2025

Documento assinado digitalmente
 FLÁVIA DANIELI MARTINS GODINHO
Data: 27/02/2025 09:00:13-0300
Verifique em <https://validar.itl.gov.br>

FLÁVIA DANIELI MARTINS GODINHO
Assistente Técnico Administrativo I da Seção Técnica de Pós-Graduação
Instituto de Biociências de Botucatu – UNESP

Instituto de Biociências - Câmpus de Botucatu -
Rua Doutor Antonio Celso Wagner Zanin, s/nº, 18618689
<https://www.ibb.unesp.br/ensino/pos-graduacao/programas-stricto-sensu/ciencias-biologicas-genetica/CNPJ:48031918002259>

Agradecimentos

Nenhuma conquista é alcançada sozinha, e esta tese não seria possível sem o apoio, incentivo e colaboração de diversas pessoas que, direta ou indiretamente, contribuíram para o desenvolvimento deste trabalho e foram fundamentais para superar os desafios e alcançar os objetivos propostos. Ao longo dessa jornada, pude contar com a orientação de mentores, a dedicação de colegas, o carinho de familiares e amigos, e o suporte de instituições que acreditaram neste projeto. A cada um deles, dedico minha sinceríssima gratidão:

Agradeço ao meu orientador, Prof. Claudio de Oliveira, por aceitar me orientar, por sua constante presença, dedicação e por não me expulsar do laboratório desde quando entrei na Iniciação Científica em 2017.

Agradeço aos meus familiares, em especial à minha mãe que foi a minha grande apoiadora, minha maior inspiração de mulher e de mãe, ao meu pai e meu irmão, cheios de carinho para dar, às minhas tias que foram minhas apoiadoras desde criança, minha vó que até hoje se orgulha em falar que estudo na UNESP e ao meu vô foi um dos meus primeiros incentivadores brincando comigo de pintar. Sem vocês não sou nada.

Agradeço com carinho ao meu parceiro, Renato, por todos os momentos e ensinamentos compartilhados, todas as vezes que um segurou a mão do outro enquanto o outro estava precisando, por todos os momentos passados e àqueles que estão por vir.

Aos meus queridos amigos, que compartilharam comigo ensinamentos e boas risadas em tempos tão difíceis: principalmente à Raquel e à Pamela, que me lembraram o que é ter um ponto de apoio diário e não apenas uma amizade, obrigada por me trazer de volta à vida, e ao meu querido grupo com quem compartilho momentos de trabalho e de lazer Aline Rodrigues, Victória Dandara e Mariana Kuranaka

Ao Pós-doc do meu laboratório Bruno Ferrete que se dispôs a contribuir com esse trabalho e mais ainda a me manter calma me passando confiança e tranquilidade, muito obrigada!

Agradeço a todos, sem exceção, do Laboratório de Biologia e Genética de Peixes, que também contribuíram significativamente com aprendizados acadêmicos e de vida durante esses anos de mestrado e de IC.

Agradeço ao Instituto de Biociências de Botucatu, ao programa de Pós-graduação em Genética e aos funcionários do Departamento de Biologia Estrutural e Funcional e da seção técnica de Pós-graduação, cujo trabalho permitiu a realização deste estudo. Agradeço também as agências de fomento, CAPES, FAPESP e CNPq.

Em memória a minha amada avó Dirce, que sempre acreditou nos netos e foi marcante em minha entrada na UNESP quando eu afirmei que entrei na Universidade pública ao dizer “ah, sua maledeta (com o perdão da palavra), ela disse mesmo que ia passar”. E continuo passando.

Esta pesquisa foi financiada com recursos da Coordenação de
Aperfeiçoamento de Pessoal de Nível Superior (CAPES), da Fundação de Amparo
à Pesquisa do Estado de São Paulo (FAPESP) e Conselho Nacional de
Desenvolvimento Científico e Tecnológico (CNPq)



RESUMO

O rápido desenvolvimento das tecnologias de sequenciamento de nova geração (NGS) e a redução dos custos do sequenciamento tem mudado o cenário da pesquisa biológica, possibilitando o aumento do número de genomas estudados e a proposta de grandes projetos visando o sequenciamento de muitos organismos, incluindo os não-modelos, o que pode gerar dados importantes para a conservação das espécies. O Brasil desponta como futuro líder nessa área visto que é o país com maior biodiversidade do planeta. Uma das famílias de peixes com maior número de espécies é a antiga Characidae e, dentro dela, se encontrava o gênero *Spintherobolus* onde a espécie *Spintherobolus papilliferus* é sua espécie tipo, uma espécie rara, basal, endêmica de cabeceiras de rios de Mata Atlântica, próximos à cidade de São Paulo e está em perigo (EN) de acordo com os critérios de avaliação do Livro Vermelho da Fauna Brasileira Ameaçada de Extinção (ICMBio, 2022) e pela Red List of Threatened Species (IUCN). Esse gênero, e o gênero *Amazonspinther* (monotípico), formam uma nova família Spintherobolidae que é grupo irmão de todas as demais. Além de não haver informação citogenética, nosso conhecimento atual diz respeito basicamente aos estudos de taxonomia e sistemática do grupo, e, conseqüentemente, com a posição basal dessa espécie, *S. papilliferus* se torna um interessante modelo para estudos evolutivos, já que uma abordagem cromossômica e genômica podem proporcionar um melhor entendimento para que possamos começar a desvendar os processos de evolução de grandes grupos de peixes. Assim, o presente projeto foi elaborado considerando nosso engajamento no projeto global de sequenciamento de genomas de vertebrados e visando a dificuldade em se obter preparações cromossômicas, o presente estudo tem como objetivo geral: construir um genoma de referência de alta qualidade para *S. papilliferus*, utilizando plataformas modernas de sequenciamento de genoma e determinar citogeneticamente suas características cariotípicas. Para isso, foram coletados 20 indivíduos e, através de análises citogenéticas, identificamos o número diploide $2n=50$ cromossomos, bandas de heterocromatina constitutiva em regiões pericentroméricas e intersticiais e uma marcação de regiões organizadoras de nucléolos (RONs), já as marcações de FISH para o gene DNAr 18S e 5S evidenciaram 2 e 1 pares, respectivamente, sendo que as marcações teloméricas foram observadas na região terminal de todos os cromossomos. Após o processo de formação de *contigs* e polimento, o draft genoma final foi de 922,5Mb, ou 0,9Gb. As métricas de qualidade Q30 se manteve maior que 90%, com N50 de 10.410. Todos resultados citogenéticos encontrados são novidade na literatura, dessa forma, os resultados obtidos com a realização deste estudo permitirão avançar nas caracterizações citogenéticas de *S. papilliferus* e a montagem do draft genoma é um primeiro passo para a obtenção de genoma de referência e o

primeiro genoma montado para uma espécie da antiga família Characidae com uma posição filogenética importante, utilizando uma plataforma de sequenciamento de leitura curta, realizada a partir de DNA genômico extraído de um único indivíduo.

Palavras – chaves: Ciências Biológicas, Peixes de água doce, Citogenética animal, Bioinformática.

ABSTRACT

The rapid development of next-generation sequencing (NGS) technologies and the reduction in sequencing costs have transformed the landscape of biological research, enabling an increase in the number of genomes studied and the proposal of large-scale projects aimed at sequencing numerous organisms, including non-model species. This can generate valuable data for species conservation. Brazil stands out as a future leader in this field, given its status as the most biodiverse country on the planet. One of the fish families with the highest number of species is the former Characidae, within which lies the genus *Spintherobolus*, with *Spintherobolus papilliferus* as its type species. This rare, basal species is endemic to headwaters of Atlantic Forest rivers near São Paulo and is classified as Endangered (EN) according to the criteria of the Red Book of Brazilian Fauna Threatened with Extinction (ICMBio, 2022) and the IUCN Red List of Threatened Species. This genus, along with the monotypic genus *Amazonspinther*, forms the new family Spintherobolidae, which is the sister group to all other families. Beyond the lack of cytogenetic information, our current knowledge is primarily limited to taxonomic and systematic studies of the group. Consequently, given the basal position of this species, *S. papilliferus* becomes an interesting model for evolutionary studies, as chromosomal and genomic approaches can provide a better understanding of the evolutionary processes of large fish groups. Thus, this project was developed considering our involvement in the global vertebrate genome sequencing project and aiming to address the challenges in obtaining chromosomal preparations. The general objective of this study is to construct a high-quality reference genome for *S. papilliferus* using modern genome sequencing platforms and to cytogenetically determine its karyotypic characteristics. For this purpose, 20 individuals were collected, and through cytogenetic analyses, we identified a diploid number of $2n=50$ chromosomes, constitutive heterochromatin bands in pericentromeric and interstitial regions, and nucleolar organizer regions (NORs). FISH markers for the 18S and 5S rDNA genes revealed 2 and 1 pairs, respectively, while telomeric markers were observed at the terminal regions of all chromosomes. After the contig assembly and polishing process, the final draft genome was 922.5 Mb, or 0.9 Gb. Quality metrics remained above 90% for Q30, with an N50 of 10,410. All cytogenetic results are novel in the literature, and the findings from this study will advance the cytogenetic characterization of *S. papilliferus*. The assembly of the draft genome represents a first step toward obtaining a reference genome and is the first genome assembled for a species of the former Characidae family with an important phylogenetic

position, using a short-read sequencing platform performed on genomic DNA extracted from a single individual.

Keywords: Biological Sciences, Freshwater fish, Animal cytogenetics, Bioinformatics.

Sumário

1. INTRODUÇÃO.....	12
1.1. Sequenciamento de genoma completo.....	12
1.2 Importância do genoma à nível cromossômico em <i>Spintherobolus papilliferus</i> ...	15
2. OBJETIVOS.....	19
3. MATERIAL E MÉTODOS.....	19
3.1. Material.....	19
3.2. Métodos.....	21
3.2.1. Obtenção dos tecidos para análise de dados genômicos.....	21
3.2.2. Sequenciamento de novo Illumina e montagem do draft genoma.....	22
3.2.3. Montagem do draft genoma ou genoma rascunho.....	22
3.2.4. Obtenção de cromossomos mitóticos.....	23
3.2.5. Coloração convencional com Giemsa.....	24
3.2.6. Regiões organizadoras de Nucléolo.....	24
3.2.7. Bandamento C.....	24
3.2.8. Hibridização in situ fluorescente (FISH).....	25
4. RESULTADOS E DISCUSSÕES.....	26
4.1. Coleta de tecido.....	26
4.2. Sequenciamento de novo e montagem do draft genoma.....	26
4.3. Identificação sexual e estágio de desenvolvimento dos espécimes.....	29
4.4. Análise cariotípica e citogenética.....	29
5. CONCLUSÕES.....	36
6. DIFICULDADES ENCONTRADAS.....	37
7. REFERÊNCIAS.....	38

1. INTRODUÇÃO

1.1. Sequenciamento de genoma completo

O sequenciamento por leituras curtas ou *short reads* constitui a segunda geração das estratégias de sequenciamento, sucedendo o método de Sanger e sendo amplamente conhecido como "sequenciamento de próxima geração" (Next-Generation Sequencing, NGS), mesmo no contexto atual da terceira geração de sequenciamento (Slatko, Gardner & Ausubel, 2018).

As novas tecnologias de sequenciamento reduziram significativamente os custos de sequenciamento e ampliaram a geração de novos dados de sequência, resultando no sequenciamento de um número crescente de espécies não modelo. Esse avanço tem beneficiado organismos para os quais havia pouca ou nenhuma informação genômica disponível até recentemente (Ekblom & Wolf, 2014). O NGS tem sido amplamente aplicado para a geração de montagens *de novo* em espécies que não possuíam um genoma de referência prévio, permitindo avanços substanciais no conhecimento de sequências repetitivas em diversos organismos, incluindo peixes (Utsunomiya et al., 2019; Silva et al., 2021).

O processo de sequenciamento de DNA envolve a fragmentação do material genético em pequenos pedaços, aos quais são adicionadas sequências adaptadoras nas extremidades 5' para facilitar a ligação a oligonucleotídeos dos sequenciadores. Além disso, índices moleculares ou "códigos de barras" são inseridos para rastrear a origem de cada leitura, permitindo sequenciamento multiplexado. As leituras geradas, com comprimentos que variam de 76 a mais de 1 Kb, são alinhadas e sobrepostas por algoritmos de montagem, formando contigs. A cobertura genômica, ou seja, o número de vezes que uma região é lida, aumenta a precisão da sequência. O uso de leituras pareadas (*paired-end reads*) torna a montagem mais eficiente. Em plataformas como as da Illumina, fluoróforos emitem sinais luminosos após a incorporação de nucleotídeos, permitindo a identificação precisa das sequências.

Entretanto, as leituras curtas apresentam limitações na resolução de regiões genômicas complexas, particularmente aquelas contendo sequências repetitivas longas, resultando em montagens incompletas e fragmentadas (Sedlazeck et al., 2018). O uso de leituras mais extensas possibilita maior sobreposição entre fragmentos, reduzindo a probabilidade de omissão de genes e sequências repetitivas (Ekblom et al., 2014). As plataformas modernas de sequenciamento oferecem a capacidade de gerar leituras mais longas, embora ainda significativamente menores que os cromossomos completos. Exemplos incluem as plataformas PacBio (Rhoads & Au, 2015) e Oxford Nanopore (Jain et al., 2018), que são frequentemente combinadas com a técnica Hi-C.

O principal desafio no sequenciamento genômico reside na montagem das sequências de modo a refletir a ordenação encontrada nos cromossomos. Para tanto, estão disponíveis tanto técnicas tradicionais de citogenética, como a preparação direta de cromossomos (limitadas a espécies de médio/grande porte), quanto abordagens indiretas, como o Hi-C, que permite inferir a conformação tridimensional da cromatina e determinar o número cromossômico. Essas abordagens têm possibilitado avanços substanciais, culminando na montagem completa do genoma humano (Nurk et al., 2022).

A aplicação de novas abordagens para a reconstrução genômica tem resultado em diversas descobertas, especialmente em genomas de grande porte. Um genoma bem montado é essencial para a identificação precisa de elementos estruturais e funcionais, contribuindo para análises mais acuradas das variações genéticas (Chakraborty et al., 2016). Por exemplo, estudos colaborativos identificaram aproximadamente 20.000 variantes estruturais no genoma humano, muitas das quais não haviam sido detectadas previamente com o uso exclusivo de leituras curtas (Chaisson et al., 2015; Sedlazeck et al., 2018). Em um volume especial da revista Science (edição 6648), foram apresentados genomas completos de mais de 230 espécies de primatas, abordando aspectos como padrões de especiação, impacto da hibridização na diversificação, adaptações ao frio na evolução da estrutura social e o uso de genomas para caracterizar mutações raras associadas ao risco de doenças humanas (Vignieri, 2023).

O genoma pode ser montado de duas maneiras: *de novo* e ressequenciamento. O termo sequenciamento "de novo" vem do latim e significa "desde o princípio". Refere-se a métodos utilizados para determinar a sequência de DNA quando não há nenhuma sequência genômica conhecida anteriormente e disponível para uso como referência. Os genomas ressequenciados utilizam um genoma já existente de referência, sendo conjuntos altamente contíguos, precisos e anotados com o conteúdo de seus genes e regiões repetitivas, o que aprimora significativamente os estudos genômicos, tanto em termos experimentais quanto analíticos (Brandies et al., 2019; Rhie et al., 2021).

Um genoma de referência é uma representação pontual da estrutura e organização do genoma de uma espécie. Da mesma forma que para os espécimes tipo na taxonomia, os genomas de referência estabelecem um padrão para estudos genômicos subsequentes (Ballouz et al., 2019), além disso, permitem uma caracterização genômica com foco na conservação da diversidade taxonômica e funcional em ecossistemas (Gelsinger, 2020).

Avaliar a precisão da montagem *de novo* é um desafio na bioinformática. O valor N50 é usado para estimar o tamanho dos contigs, mas um N50 alto não garante uma montagem precisa. Para verificar a qualidade, alinhamentos com sequências de referência, como BLAST,

são utilizados para identificar erros e fragmentações. Os montadores de novo utilizam dois principais algoritmos: OLC, ideal para genomas menores e leituras longas, e grafos de Bruijn, mais eficiente para grandes volumes de leituras curtas. No método de Bruijn, as leituras são fragmentadas em k-mers e conectadas por sobreposição para formar contigs. A escolha do valor de k é crucial, pois valores menores podem gerar erros, enquanto valores maiores dificultam a montagem em regiões repetitivas. Estratégias iterativas que combinam diferentes k têm sido exploradas para melhorar a precisão da montagem.

Grandes iniciativas foram estabelecidas, como por exemplo o Projeto Genoma 10K (G-10K), com o propósito de fornecer uma grande base de genomas de alta qualidade para vertebrados de referência. Essa base de dados serviria como fundamento para inúmeras pesquisas futuras. Uma das principais metas do G-10K é o sequenciamento de 10 mil espécies de vertebrados (<https://vertebrategenomesproject.org/>). Contudo, diversos desafios precisam ser superados para que essa meta seja alcançada, tanto no processo de montagem dos genomas (Genome 10k Community of Scientists 2009; Koepfli *et al.*, 2015), como na obtenção do DNA de espécies raras ou de difícil prospecção.

O objetivo final de qualquer projeto de genômico atual é montar um genoma sem intervalos e resolvido por haplótipos, o que possibilitará aos pesquisadores investigarem a variação estrutural (SV) que, de outra forma, seria perdida com sequenciamento de leitura curta (Phillippy 2017).

O acesso a um genoma montado para uma espécie de interesse pode contribuir para o avanço do conhecimento em diversos aspectos biológicos, tais como a ecologia (Ekblom e Wolf, 2014), biologia evolutiva e especiação (Ellegren, 2014). É importante que haja representação em todas as linhagens para viabilizar estudos comparativos da biologia genômica, os quais, por sua própria natureza, dependem do conhecimento das posições de suas sequências ortólogas (Ruiz-Herrera e Farré-Belmonte, 2021).

O primeiro peixe a ter seu genoma totalmente sequenciado foi um baiacu, *Takifugu rubripes*, uma espécie de grande importância para a pesca e aquicultura, e significativa como modelo de vertebrado com o menor genoma conhecido (Aparicio *et al.*, 2002). O uso de novas tecnologias vem aperfeiçoando esse genoma inicial, resultando hoje em uma versão altamente precisa do genoma completo da espécie (<https://www.ncbi.nlm.nih.gov/genome/?term=Takifugu+rubripes>). Uma busca recente no portal do Centro Nacional de Biotecnologia dos EUA Informação Genoma NCBI (<https://www.ncbi.nlm.nih.gov/genome/>), revelou informações disponíveis sobre o genoma de 1058 espécies de peixes, embora sem especificação do nível de montagem de cada um desses

genomas. Até o momento, o único genoma completo publicado para Characidae é o de *Astyanax mexicanus* (Warren *et al.*, 2021). Neste estudo, os autores descrevem o genoma da espécie e conduzem uma comparação entre peixes da superfície e populações de caverna, sugerindo, por meio de análises de mapeamento de loci quantitativos (QTL), a existência de genes candidatos à perda ocular, como o *dusp26*.

1.2 Importância do genoma a nível cromossômico em *Spintherobolus papilliferus*

Characidae foi por muito tempo considerada a maior família de peixes de água doce neotropicais, abrangendo mais de mil espécies, o que correspondia a mais da metade das espécies viventes de Characiformes. Ao longo dos anos diversos estudos filogenéticos (Oliveira *et al.*, 2011; Mirande, 2019; Betancur-R *et al.*, 2019; Melo *et al.*, 2021) sugeriam que Characidae poderia ser subdividida em quatro linhagens principais: 1) Spintherobolinae [*Spintherobolus* e o gênero monotípico *Amazonspinther*]; 2) o clado A ou Stevardiinae; 3) o clado B, que inclui Aphyocharacinae, Characinae, Cheirodontinae, Exodontinae e Tetragonopterinae; e 4) o clado C ou Stethaprioninae, composto pelos antigos Stethaprioninae, Rhoadsiinae e inúmeros grupos sem nome. Existiam diversos aspectos interessantes a serem investigados em Characidae, incluindo o fato de que, enquanto a subfamília Spintherobolinae apresentava apenas cinco espécies conhecidas, a subfamília Stethaprioninae (clado C) possuía mais de 650 (Eschmeyer e Fong, 2024).

A análise filogenética mais recente de Melo *et al.* (2024), baseada em elementos ultraconservados (UCEs), levou a uma reclassificação taxonômica. O clado até então conhecido como Characidae deu origem a quatro famílias: Spintherobolidae (*Amazonspinther* + *Spintherobolus*; = Spintherobolinae de Mirande, 2018); Stevardiidae (=clado A de Malabarba & Weitzman, 2003; Stevardiinae de Mirande, 2010, Oliveira *et al.*, 2011, Thomaz *et al.*, 2015, etc.); Acestrorhamphidae (=Stethaprioninae de Mirande, 2018); e Characidae, agora restrita às subfamílias Aphyocharacinae, Cheirodontinae, Exodontinae, Tetragonopterinae e Characinae. (Figura 1).

Dentre as famílias destaca-se Spintherobolidae recuperada como grupo-irmão do clado maior composto por Stevardiidae, Characidae e Acestrorhamphidae (Melo *et al.*, 2024), contrariando hipóteses anteriores que sugeriam sua proximidade com Cheirodontinae (Bühnheim *et al.* 2008). A monofilia da família é sustentada por 15 sinapomorfias, sendo a ausência do mesocoracóide uma das principais características compartilhadas. A família representa o grupo menos diverso entre os caracídeos com apenas cinco espécies, quatro delas descritas em *Spintherobolus*.

Este gênero inclui a espécie *Spintherobolus papilliferus* (Weitzman & Malabarba, 1999) (Figura 2), uma espécie endêmica da Mata Atlântica, classificada como em perigo de extinção (EN) de acordo com os critérios de avaliação do Livro Vermelho da Fauna Brasileira Ameaçada de Extinção (ICMBio, 2022) e pela Lista Vermelha de Espécies Ameaçadas da (IUCN). As principais ameaças incluem alterações no habitat, expansão das áreas urbanas e poluição. Entre as espécies de *Spintherobolus*, *S. papilliferus* ocupa uma posição basal, enquanto *S. broccae* forma um grupo-irmão com *S. ankoseion* e *S. leptoura*. Estudos baseados em loci UCE confirmam a monofilia do grupo e sugerem um histórico de vicariância que pode ter influenciado a distribuição geográfica das espécies (Melo et al., 2024). Além do exposto acima, *S. papilliferus* é a espécie tipo do gênero *Spintherobolus* e da família Spintherobolidae.

Estudos citogenéticos em peixes revelavam uma grande diversidade de cariótipos entre espécies e populações, com número diplóide modal igual a $2n = 50-52$ e número fundamental variando de 56 a 132 (Arai, 2011), indicando uma ampla variação na estrutura do cariótipo, apesar de os números de cromossomos diplóides serem relativamente constantes. Ao correlacionarmos informações sobre a estrutura cariotípica com os grupos monofiléticos na antiga Characidae, observávamos que uma característica marcante na antiga subfamília Stethaprioninae (clado C) era a presença, em todas as espécies analisadas, de um grande par metacêntrico (par 1), como recentemente confirmado para Rhoadsiinae (Sánchez-Romero et al., 2015) e $2n=50$ cromossomos. Esse par cromossômico característico não é observado na antiga subfamília Stevardiinae (Guimarães et al., 1995) nem no Clado B (Souza et al., 1995), e, além disso, as mesmas apresentam $2n=52$ cromossomos. Para a antiga linhagem Spintherobolinae não havia informação citogenética disponível.

Considerando a posição basal da linhagem Spintherobolinae na filogenia da antiga Characidae (Melo et al., 2021), *Spintherobolus* emerge como um modelo interessante para estudos evolutivos, visto que uma abordagem cromossômica e genômica pode proporcionar um melhor entendimento dos processos em evolução de grandes grupos de peixes.

Segundo Melo et al. (2024), algumas características são identificadas como sinapomorfias, ou seja, traços derivados que são compartilhados por membros de um clado e que podem ser usados como evidências para a definição e diferenciação desses grupos:

- 1-Ausência do osso mesocoracóide em Spintherobolidae;
- 2-Presença de quatro dentes pré-maxilares e oito raios branqueados na nadadeira dorsal em Stevardiidae;
- 3-O maior par cromossômico metacêntrico como possível sinapomorfia para Acestrorhamphidae.

Sánchez-Romero et al. (2015) hipotetizaram que o grande par de cromossomos metacêntrico encontrado apenas na antiga Stethaprioninae, e agora Acestrorhamphidae, poderia representar uma característica única e derivada (apomorfia) da subfamília. Isso poderia por um lado, reforçar sua monofilia e, por outro lado fornecer um ponto de partida interessante para investigar as causas da ampla biodiversidade encontrada no grupo.

Com a separação da Spintherobolinae em uma família distinta, *S. papilliferus* antes relacionado a questões de diversidade genética e citogenética da família Characidae, agora precisa ser abordado dentro de Spintherobolidae, considerando suas relações filogenéticas e a reinterpretação de seus aspectos evolutivos, como estrutura cromossômica e genômica. Portanto, a integração de dados genômicos para esse grupo passa a ser uma prioridade.

Assim, a estreita ligação entre DNA e a estrutura cromossômica destaca a importância de integrar dados citogenéticos e genômicos para uma compreensão mais abrangente. Essa abordagem integrada não apenas fornece uma fonte mais completa de informações, mas também é fundamental para a compreensão da evolução e adaptação molecular.

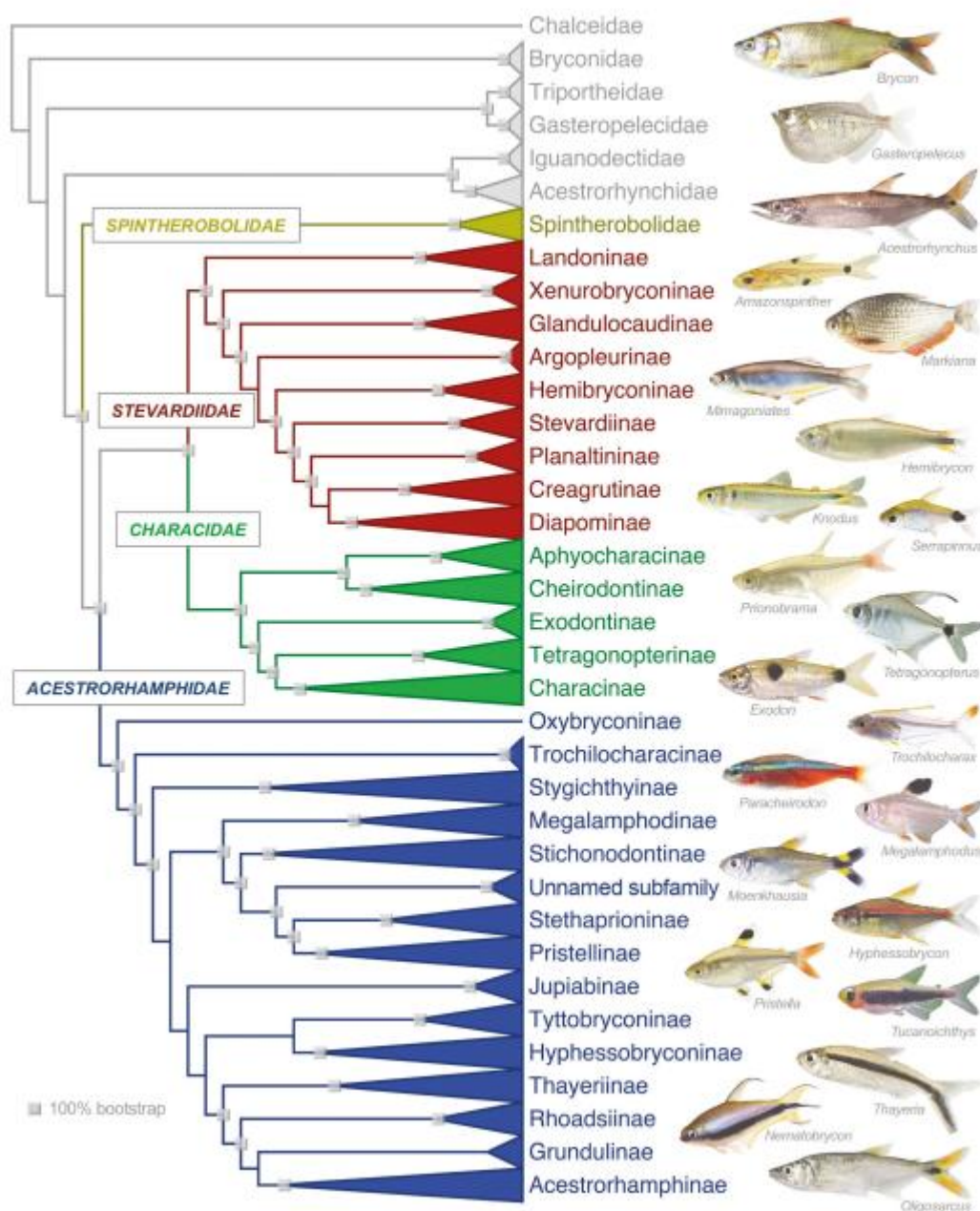


Figura 1: Árvore filogenética baseada em elementos ultraconservados (UCEs), adaptada de Melo et al. (2024), mostrando a reclassificação do grupo Characidae em quatro famílias distintas, com *Spintherobolus* na base da árvore. Observe a posição basal de *Spintherobolidae*.



Figura 2: Exemplar de *S. papilliferus*, foto por Bruno Melo.

2. OBJETIVOS

O presente projeto propõe-se a cariotipar e gerar o primeiro genoma de ampla cobertura para *Sphinterobolus papilliferus*, com o objetivo de criar um banco de dados para futuros estudos genômicos na família Characidae, contribuindo para posteriores comparações e discussões sobre a conservação do gênero *Spintherobolus* e de peixes da antiga família Characidae.

Objetivos específicos:

- (1) Caracterizar o cariótipo de *Sphinterobolus papilliferus*;
- (2) Obter a montagem do genoma dessa espécie;

3. MATERIAL E MÉTODOS

3.1. Material

Foram feitas duas coletas com o uso de redes de puçás. Uma coleta em janeiro de 2024 e outra em julho de 2024, totalizando 20 exemplares da espécie *S. papilliferus*, de um afluente do rio Juquiá (Figura 3), rio Ribeira de Iguape, Juquitiba, São Paulo (23°59'13.20"S 47°00'10.8"N). A inclusão de *S. papilliferus* na lista das espécies a serem sequenciadas pelo Projeto Genoma dos Vertebrados (VGP - *Vertebrate Genomes Project*) e sua presença na lista da IUCN como em estado de perigo, como já mencionado, reforçam a importância e a relevância da escolha desta espécie. Os espécimes foram coletados na natureza, mediante licença de coleta do IBAMA/SISBIO.

Os espécimes foram separados de acordo com o sexo, mediante análise morfológica de tecido gonadal. Amostras de tecidos incluindo cérebro, coração, brânquias, fígado, gônadas e músculo, foram coletadas para utilização em diferentes técnicas atuais e futuras a este projeto (Tabela 1).

Os exemplares utilizados na amostragem foram devidamente identificados. Quatro indivíduos foram fixados em formol 10% e, em seguida, conservados em álcool etílico 70% para preservar a suas características morfológicas. Os restantes dos exemplares foram conservados em álcool 100% para preservar o DNA. Posteriormente, os espécimes foram depositados na coleção de peixes do Laboratório de Biologia e Genética de Peixes do Departamento de Biologia Estrutural e Funcional, IB-UNESP, Botucatu, SP.

Tabela 1: Relação dos tecidos congelados e suas aplicações para as técnicas.

Tecido	Brânquia	Fígado	Coração	Músculo	Cérebro	Gônada	Rim
Destino	Seq. Short reads Cultura celular	Seq. Long read - PacBio	Seq. Long read - PacBio	Seq. Short reads Hi-C Cultura celular	Seq. Shorts reads Hi-C Cultura celular	Short read - Transcri ptoma	Preparaç ão cromossô mica

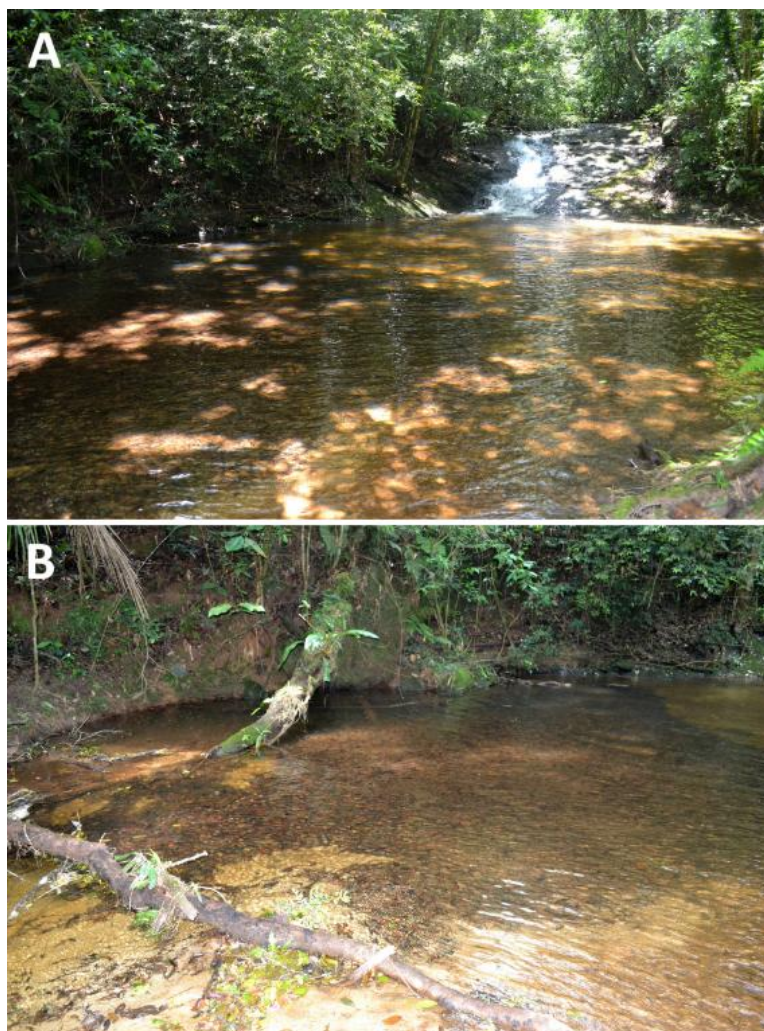


Figura 3: Adaptado de Mattox *et al.* (2023), trecho de afluente do rio Juquiá, rio Ribeira de Iguape, município de Juitiba, Estado de São Paulo, onde os espécimes de *Spintherobolus papilliferus* foram coletados para este estudo. A. Vista geral mostrando as corredeiras ao fundo. B. Piscina de corredeiras onde os espécimes foram capturados.

3.2. Métodos

3.2.1. Obtenção dos tecidos para análise de dados genômicos

Uma montagem de genoma completo ocorre pela combinação dos fragmentos do DNA de *shorts reads* gerados pela plataforma de sequenciamento Illumina, que futuramente dará suporte aos *longs reads*, gerados pelo sequenciamento PacBio, até atingir uma escala cromossômica com o auxílio de *Hi-C reads*. Para isso, as amostras de tecidos foram obtidas de exemplares recém coletados e congelados em nitrogênio líquido para a preservação máxima do DNA. Os tecidos foram então armazenados em criotubos e mantidos em ultra freezer a -80°C , conforme as orientações do Projeto Genoma dos Vertebrados (VGP <https://vertebrategenomesproject.org/technology>) até o momento da realização das suas respectivas técnicas, mencionadas anteriormente (Tabela 1).

3.2.2. Sequenciamento *de novo* Illumina e montagem do draft genoma pela Illumina

Para uma montagem do genoma de referência ideal, são utilizados as *longs reads* que, usando softwares adequados, geram os *contigs*, que serão ordenados e orientados em *scaffolds* usando os *Hi-C reads* finalizando com as *shorts reads*, pela Illumina, para resultar em uma alta qualidade da montagem sem erros e detectando variações como SNP e INDELS.

O DNA genômico de um indivíduo de *S. papilliferus* foi extraído utilizando o Wizard Genomic DNA Purification Kit (Promega), seguindo as instruções fornecidas pelo fabricante, e sua concentração foi estimada utilizando um fluorímetro Qubit 2.0 (Invitrogen™). A amostra deve apresentar uma concentração mínima de 100 ng/μL, com alta qualidade e peso molecular para ser destinada ao sequenciamento.

O DNA extraído foi de um indivíduo depositado na coleção do nosso laboratório, LPB20540, de uma população do rio Itapanhaú, Bertioga, São Paulo. A extração ocorreu antes de realizarmos as coletas para este projeto pois surgiu uma oportunidade de sequenciamento, e ainda não tínhamos as amostras coletadas. Essas sequências poderão ser usadas para a montagem do genoma, pois segundo Matto'x *et al.* (2023), não há divergência genética entre as duas populações. Também podemos usar para comparação com a amostra da população de Jiquitiba.

Para este sequenciamento utilizamos a plataforma NovaSeq. 600 System com leituras paired-end 2x150bp, feita pela empresa Daicel Arbor Biosciences, localizadas nos Estados Unidos, com cobertura de sequenciamento de aproximadamente 30x, considerando que o genoma de *S. papilliferus* tem tamanho estimado em 1.87 Gb pelo portal *Genomes on a Tree* (GoaT) (<https://goat.genomehubs.org/>).

3.2.3. Montagem do draft genoma ou genoma rascunho

As reads brutas passaram por uma etapa de filtragem de qualidade e correção no tamanho, utilizando o software fastp (Chen, 2023) e as informações fornecidas pela plataforma Illumina de adaptadores e tags, para a remoção das mesmas (Tabela 2). A construção do draft genoma foi feita com o software gatb-minia-pipeline, sendo um programa multi-k buscando alta contiguidade, ele utiliza uma sequência de programas, sendo eles Minia 3 (<https://github.com/GATB/gatb-minia-pipeline>) para montagem. A sequência rascunho do genoma foi submetida a um polimento com o software DeepVariant v1.5.0 (<https://github.com/google/deepvariant>), uma pipeline de análise que usa uma rede neural

profunda para chamar variantes genéticas de dados de sequenciamento de DNA de próxima geração. Não foi possível realizar o processo de *scaffolding* para esse momento do projeto, por se utilizar mais memória RAM e tempo do que temos disponível.

As estatísticas do genoma final serão calculadas pelo QUAST v5.0.2 (Gurevich et al., 2013) e uma análise de integridade do conjunto de genes será realizada através do software BUSCO v5.2.2 (Simão et al., 2015; Waterhouse et al., 2018) com o banco de dados fornecido para genes ortólogos da base de dados para Actinopterygii (Actinopterygii_odb10).

Tabela 2: informações dos adaptadores

Adaptadores
i7:GATCGGAAGAGCACACGTCTGAACTCCAGTCAC*ATCTCGTATGCCGTC TTCTGCTTG
i5:AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT*GTGTAGATTCCGGT GGTCGCCGTATCATT
Sequências de tags
i7-A_002:GCGCCGTG
i5-B_002:CGATGTTT

3.2.4. Obtenção de cromossomos mitóticos

Aproveitamos a coleta dos tecidos dos espécimes para realizar uma preparação direta de cromossomos metafásicos mitóticos utilizando a técnica proposta por Foresti *et al.* (1989) com algumas adaptações. Começamos retirando o tecido de interesse, realizando a digestão mecânica e acrescentando a colchicina de duas formas em diferentes tempos e concentrações, para a inibição da polimerização dos microtúbulos. Posteriormente, realizamos a hipotonização das células em suspensão deixando agir por 20 minutos à 37°C. Após, fizemos a fixação celular pela mistura de metanol/ácido acético (fixador) e, ao final, de 200 a 700µl de fixador foi adicionado e transferido para um *eppendorf* e mantido em freezer a -20°C.

Testamos em Meio Hanks por 15 minutos com a concentração da colchicina à 0,005%, 0,0025% e 0,0017% e por injeção a 40 minutos.

As lâminas foram analisadas em fotomicroscópio óptico de fluorescência (Olympus BX61) e as imagens foram capturadas utilizando o programa CellSens Dimension (Olympus).

A composição final das imagens foi realizada utilizando o software editor de imagem Adobe Photoshop CS6.

3.2.5. Coloração convencional com Giemsa

Aproximadamente 15 µL da suspensão celular foi gotejada em uma lâmina limpa apoiada em vapor de água e secada em estufa a 37°C. Então, os cromossomos foram corados em solução de Giemsa a 5% em tampão fosfato (pH=6,8) por 8 minutos. Os cromossomos foram observados em fotomicroscópio óptico e, finalmente, os cromossomos foram arranjados de acordo com os tipos metacêntricos (m), submetacêntricos (sm), subtelocêntricos (st) e acrocêntricos (a), seguindo a classificação estabelecida por Levan *et al.* (1964), emparelhados com seus prováveis homólogos e organizados em ordem decrescente de tamanho, para a disposição final do cariótipo.

3.2.6. Regiões organizadoras de Nucléolo

Para identificar as regiões organizadoras de nucléolo, foi utilizada a técnica de impregnação de nitrato de prata descrita por Howell e Black (1980), com algumas modificações. As lâminas foram hidrolisadas em HCl 1N por 10 segundos e então lavadas em água destilada. Posteriormente, foram adicionadas três gotas de uma solução coloidal reveladora juntamente com seis gotas de uma solução de nitrato de prata 50%, então as lâminas foram cobertas com uma lamínula e mantido em banho maria à 60°C aproximadamente por dois minutos ou até adquirir uma coloração marrom/dourada. As lâminas foram novamente lavadas em água destilada, secadas em estufa a 37°C e observadas em fotomicroscópio óptico.

3.2.7. Bandamento C

As regiões de heterocromatina constitutiva foram evidenciadas pela técnica descrita por Sumner (1972) com modificações. As lâminas foram emergidas em uma solução de HCL 0,2N em temperatura ambiente por 30 minutos. Então, as lâminas foram colocadas numa solução de BaOH à 60°C por 15 segundos, depois foram lavadas em um jato de água corrente e imediatamente imersas na solução de HCL 1N por duas vezes a 60°C por 5 segundos cada. As lâminas foram lavadas em água corrente e colocadas em uma solução de 2xSSC à 60°C por 15 minutos. As lâminas foram coradas com Giemsa e observadas em fotomicroscópio óptico de fluorescência (Olympus BX61).

3.2.8. Hibridização *in situ* fluorescente (FISH)

A técnica de FISH foi realizada com alta estringência procedida como descrito em Pinkel *et al.* (1986), com algumas modificações estabelecidas pelo nosso laboratório. Duas sondas foram utilizadas para os experimentos de Hibridação *in situ* fluorescente: DNAr 18S e 5S através de primers descritos por Utsunomia *et al.* (2016), usando DNA genômico da própria espécie. As sondas foram produzidas e marcadas por reação em cadeia da polimerase (PCR) utilizando biotina-16-dUTP (Roche Applied Science) para a sonda DNAr 18S e digoxigenina-11-dUTP (Roche Applied Science) para a sonda DNAr 5S.

Os cromossomos metafásicos foram incubados em RNase (100 µg/ml) por 1h à 37°C, lavados em solução salina e, então, fixados em formaldeído 1%/PBS 1x/MgCl₂ 50mM por 10 minutos à temperatura ambiente. Após, a desnaturação do DNA cromossômico foi conduzida em formamida 70% à 70°C por 5 minutos e desidratada em série alcoólica 70%, 85% e 100% por 3 minutos. Já a solução de hibridização (10% sulfato dextrano, 50% formamida, 20xSSC, e 3-4 uL da sonda) foi desnaturada à 98°C por 10 minutos. O processo de hibridização foi realizado overnight em câmara úmida a 37°C.

Após a hibridização, as lâminas foram lavadas em formamida 15% por 20 minutos à 42°C e por 15 minutos em 0,1xSSC à 60°C. Posterior as lavagens, as lâminas foram incubadas em uma solução de bloqueio (2,5g de leite em pó, 40mL 20xSSC e 10mL água MilliQ) por 15 minutos em temperatura ambiente e, em seguida, a detecção do sinal das sondas foi feita utilizando anticorpos (1uL de avidina-FITC e 0,5 de antidigoxi-rodamina acrescido de 100uL da solução bloqueio por lâmina) coberto por lamínula por uma hora incubadas em câmara úmida à 37°C, então coradas com 20 µL de Vectashield® Mounting Medium com DAPI (Vector) e visualizados e capturados em fotomicroscópio óptico de fluorescência (Olympus BX61).

4. RESULTADOS E DISCUSSÕES

4.1. Coleta de tecido

Foi realizada uma coleta em janeiro de 2024 e outra em julho de 2024, totalizando 20 exemplares da espécie *S. papilliferus*. No entanto, foi possível apenas obter tecidos de 8 indivíduos (Tabela 3), pois alguns morreram durante o tempo em que permaneceram no aquário e outros foram utilizados com foco na citogenética com a injeção da colchicina e retirada dos rins.

Tabela 3: Relação entre espécimes coletados e retirada de tecidos para realização das técnicas moleculares de sequenciamento, cultura celular e obtenção de cromossomos metafásicos direta.

Voucher amostra	Sexo	Tecido
106316	Macho	Todos tecidos coletados
106317	Fêmea	Todos tecidos coletados
106318	Fêmea	Ausência de coração, fígado e gônada
106319	Macho	Todos tecidos coletados
106320	Fêmea	Ausência de coração
106321	Fêmea	Todos tecidos coletados
106322	Fêmea	Somente realizado a citogenética
106323	Fêmea	Ausência de coração
106324	Fêmea	Somente realizado a citogenética
106355	Fêmea	Somente realizado a citogenética
106356	Fêmea	Todos tecidos coletados
106357	Fêmea	Todos tecidos coletados

4.2. Sequenciamento *de novo* e montagem do draft genoma

Não foi possível realizar os sequenciamentos para a obtenção dos *Hi-C reads* e *long reads* no Instituto Max Planck pelas amostras apresentarem baixo peso molecular na extração do DNA genômico. O que influenciou no resultado foi o meio de transporte sugerido pelo responsável do Instituto, sendo o respectivo tecido armazenado em RNAlater para o sequenciamento Hi-C e DMSO + 20% NaCl saturado para o sequenciamento do genoma completo por PacBio® Sequel II System.

Os resultados brutos do sequenciamento Illumina NovaSeq e as métricas para *S. papilliferus* são apresentados em detalhes na Tabela 4.

O Fastp foi utilizado para o controle de qualidade inicial e o pré-processamento dos dados de sequenciamento. Após a filtragem, o comprimento médio das leituras geralmente diminui entre as espécies, indicando a remoção de leituras de baixa qualidade ou sequências de adaptadores. Após a filtragem melhorou a qualidade conforme evidenciado pelo aumento na porcentagem de bases Q20 e Q30.

Tabela 4: Rendimento do sequenciamento e métricas brutas da biblioteca para a espécie sequenciada utilizando a plataforma de sequenciamento de leitura curta Illumina NovaSeq.

Espécie		<i>S. papilliferus</i>
Geral	fastp versão	0.20.0 (https://github.com/OpenGene/fastp)
	Sequenciamento	2×150 bp
	Tamanho total das reads antes de filtrar	150bp, 150bp
	Tamanho total das reads depois de filtrar	149bp, 149bp
Antes de filtrar	Reads totais (M)	780 M
	Bases totais (Gb)	117 G
	Q20 bases	97,14%
	Q30 bases	92,68%
	Conteúdo GC	41,03%
After filtering	Reads totais (M)	775,78 M
	Bases totais (Gb)	115,65 G
	Q20 bases	97,56%
	Q30 bases	93,46%
	Conteúdo GC	41,25%

Após o processo de formação de *contigs* e polimento, o draft genoma final foi de 922,5Mb, ou 0,9Gb. Métricas de qualidade e estatísticas de completude são mostradas na Tabela 5. Somente foi observado uma pequena variação no tamanho total do genoma entre a primeira montagem e o último polimento, sendo de 922555381bp a 922556904bp, respectivamente.

Valores do BUSCO nos indicam que o genoma ainda está fragmentado, o que já era esperado, pois a natureza fragmentada das montagens dos genomas preliminares também pode ser atribuída às limitações da tecnologia de sequenciamento de leitura curta da Illumina em resolver regiões genômicas repetitivas e de baixa complexidade (Guiglielmoni *et al.*, 2024)

Tabela 5: Métricas de qualidade pelo software QUATS e estatísticas de completude pelo BUSCO.

Estatísticas de montagem	<i>S. papilliferus</i>
Tamanho da montagem (bp)	922556904
Tamanho da montagem (Mb)	922,5
Número de <i>contigs</i>	320.176
Valor N50 (bp)	10.410
Maior <i>contigs</i>	197.186
No. <i>contigs</i> >5.000 bp	48.548
No. <i>contigs</i> >10.000 bp	23.148
No. <i>contigs</i> >50.000 bp	784
Conteúdo GC (%)	40,52
Complete BUSCOs (%)	2112
Cópias completas e únicas BUSCOs (%)	2098
Completas e duplas BUSCOs (%)	14
Fragmentados BUSCOs (%)	711
Perdidas BUSCOs (%)	817

Estudos genômicos e filogenômicos são ferramentas essenciais para auxiliar em estudos como reconstrução de árvores evolutivas e entender as relações evolutivas entre diferentes espécies (Felsenstein, J., 2004), seleção natural e adaptação (Nielsen, R., *et al.*, 2005), genômica comparativa, comparando genomas de diferentes espécies para identificar elementos conservados e divergentes, o que pode fornecer insights sobre a evolução genômica (Ureta-Vidal, 2023), evolução de genomas em espécies ameaçadas, estudando a diversidade genética e a história evolutiva de espécies ameaçadas, auxiliando em esforços de conservação. Todas essas alternativas são importantes para o contexto de *S. papilliferus*, visto que no recente artigo publicado de Melo *et al.* (2024), a árvore evolutiva da antiga família Characidae foi reconstruída e Brandies (2019) diz que estudos comparativos envolvendo a conservação das espécies são cada vez mais necessários.

Um componente essencial para alcançar uma compreensão completa da biologia e da história evolutiva de uma espécie é possuir um genoma de referência de alta qualidade. No entanto, o maior desafio continua sendo o isolamento de DNA de alto peso molecular em organismos ao longo de todo o espectro da vida. (Lee *et al.*, 2023).

Uma compreensão futura com a futura fase de sequenciamento de leituras longas irá resolver essas problemáticas à medida que mais dados genômicos completos estiverem disponíveis a partir de estudos que se concentram em enriquecer as leituras.

4.3. Identificação sexual e estágio de desenvolvimento dos espécimes

Dos exemplares coletados, foi possível identificar 15 fêmeas e 5 machos. A determinação do sexo foi realizada por meio de análise microscópica das gônadas. Não existem cromossomos sexuais.

4.4. Análise cariotípica e citogenética

Foi realizada a preparação cromossômica em todos os espécimes, sendo que algumas foram preparadas *pós mortem*. Enquanto a maioria foi realizada após a retirada e congelamento dos tecidos (Tabela 1).

O número diplóide (Figura 4 e 11B) é de $2n=50$ cromossomos, com fórmula cariotípica de $8m+10sm+5st+2a$ e número fundamental de 94. Foi preciso vários testes com colchicina para obtermos um número satisfatório de metáfases. Para análises de blocos de heterocromatina constitutiva (Figura 5), podemos observar a marcação em 10 pares cromossômicos, localizados nas regiões pericentroméricas e intersticiais, mantendo esse padrão em outras espécies dispersas pela antiga Characidae, sugerindo um padrão de conservação.

Além disso, para as regiões organizadoras de nucléolos (Figura 6 - box) evidenciamos a marcação em uma porção terminal do par 24, um cromossomo acrocêntrico, diferente de *H. multifasciatus* (Soares *et al.*, 2021) (Acestrorhamphidae), *Serrapinnus heterodon* (Peres *et al.*, 2007) (Characidae) e *Momagoniatus microlepis* (Torres *et al.*, 2008) (Stervadiidae), com marcações em regiões centroméricas para *H. multifasciatus* e terminais para as duas últimas espécies, sendo em cromossomos metacêntricos (par 5), submetacêntricos (par 9) e metacêntricos (par 5), respectivamente.

Já as marcações de FISH para o gene DNAr 18S (biotina – verde) e 5S (digoxigenina – vermelho) evidenciaram 2 e 1 pares (Figura 7) nos pares 22 e 25, e no par 3, respectivamente, sendo que as marcações teloméricas (Figura 8) foram observadas na região terminal de todos os cromossomos, indicando que não houve um processo de fusão. Comparando o gene DNAr 18S com *H. multifasciatus* (Soares *et al.*, 2021), Acestrorhamphidae, percebemos que está evidenciado somente no par 4, sendo que para esta espécie, o DNAr 5S está presente no grande par metacêntrico, diferente de *S. papilliferus* que não possui nenhuma marcação no primeiro par.

Esse resultado citogenético demonstra relevância pois, segundo o portal *Genomes on a Tree* (GoaT), *S. papilliferus* tem como ancestral direto *Odontostilbe stenodon*, o qual possui 52 cromossomos, localizado na subfamília Cheirodontinae, Characidae, antigo clado B. Encontramos descrito na literatura o número modal para *Odontostilbe pequirá*, o qual faz parte do mesmo gênero, com o mesmo número modal (Troy, 2010) (Figura 9A). A mesma comparação também é feita com cromossomos da Stervadiidae, como por exemplo em *Momagoniatus microlepis* (Torres *et al.*, 2008), a qual também possui 52 cromossomos (Figura 9B). É importante destacar que o tamanho do primeiro par comparado ao segundo nessas duas famílias não é relativamente diferentes para chamarmos o primeiro par de grande par metacêntrico. Em comparação, espécies de Acestrorhamphidae, antiga Stethaprioninae, como por exemplo em *H. multifasciatus* (Soares *et al.*, 2021), possuem o grande par metacêntrico quando avaliamos o primeiro com o segundo par (Figura 10A).

Definir o cariótipo de *S. papilliferus* foi uma tarefa árdua, pois o nono par causou dúvidas se seria metacêntrico ou submetacêntrico e, seguindo as métricas por Levan *et al.* (1964), esse par se enquadrava no limite da medida sendo classificado como submetacêntrico, tornando o primeiro par, o grande metacêntrico (Figura 10B), o que nos leva a discutir a sinapomorfia de Acestrorhamphidae possuindo o maior par cromossômico metacêntrico e como foi a evolução cromossômica da antiga Characidae.

Ademais, essas informações endossam trabalhos anteriores como quando Mello *et al.* (2022) discutem a separação da proximidade entre *S. papilliferus* e o seu antigo grupo irmão Cheirodontidae e demonstram que, atualmente, é grupo irmão de *Amazonspinther* para a qual não há dados citogenéticos e (2) dados ainda mais recentes de Melo *et al.* (2024), como já comentado anteriormente, separam a antiga linhagem Spintherobolinae em uma nova família Spintherobolidae.

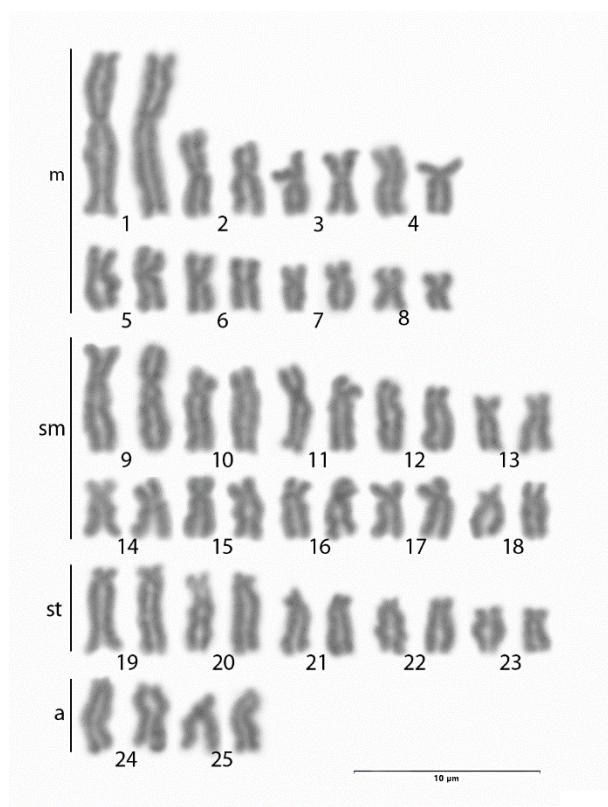


Figura 4: Cariótipo de *Spintherobolus papilliferus*, $2n=50$ cromossomos com $6m+12sm+26st+6a$. Barra = 10 μm.

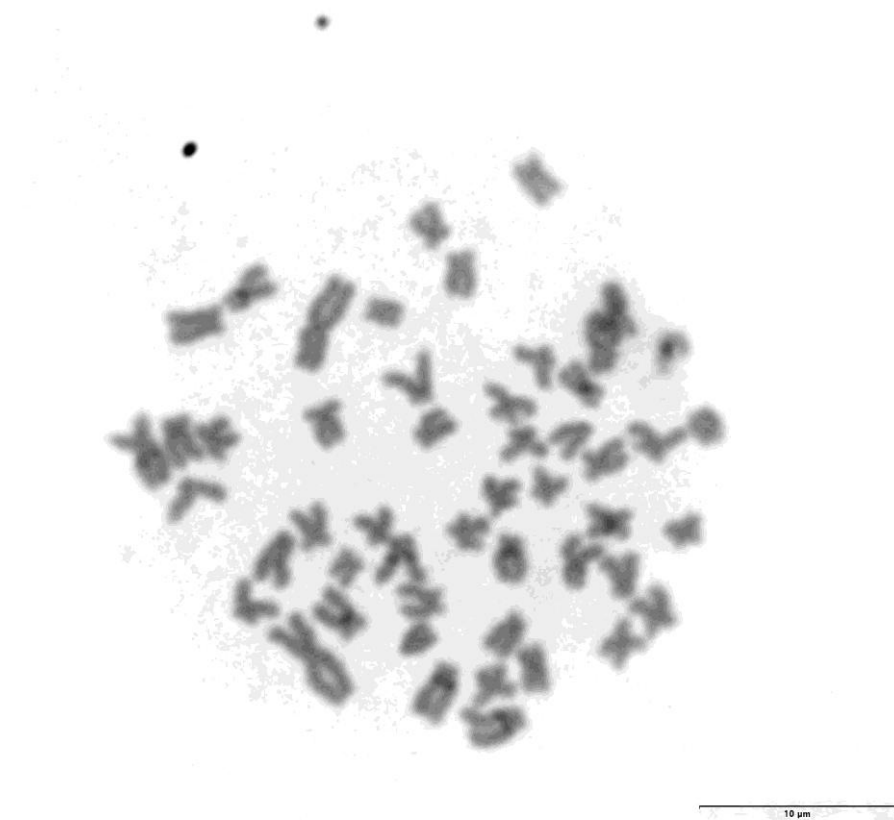


Figura 5: Cromossomos metafásicos de *S. papilliferus* após Bandamento C. Barra = 10 μm.

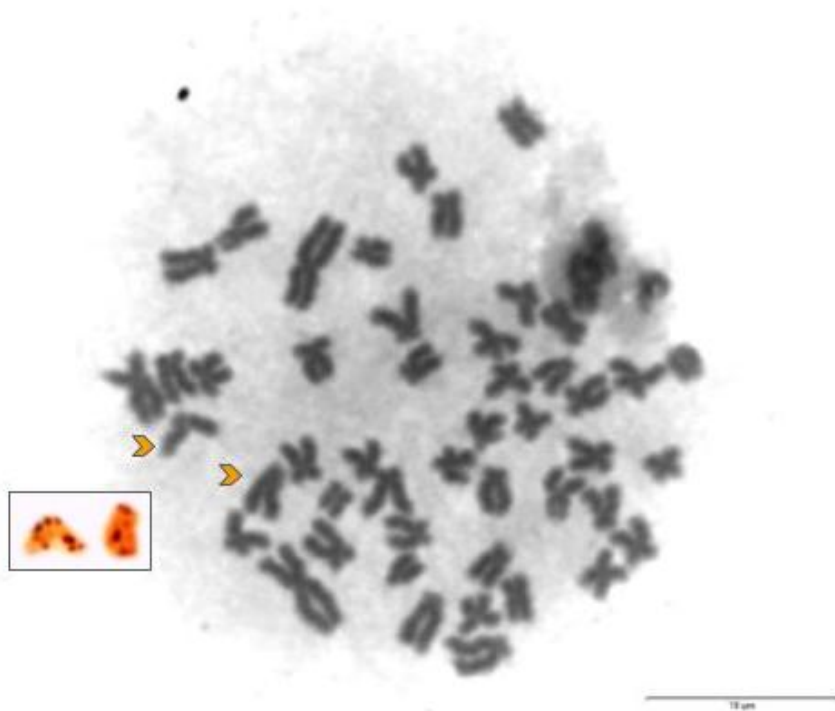


Figura 6: Cromossomos metafásicos de *S. Papilliferus*, $2n=50$. A região organizadora de nucléolo está presente na caixa, sendo as pontas de setas os cromossomos correspondentes na metáfase. Barra = 10 μm .

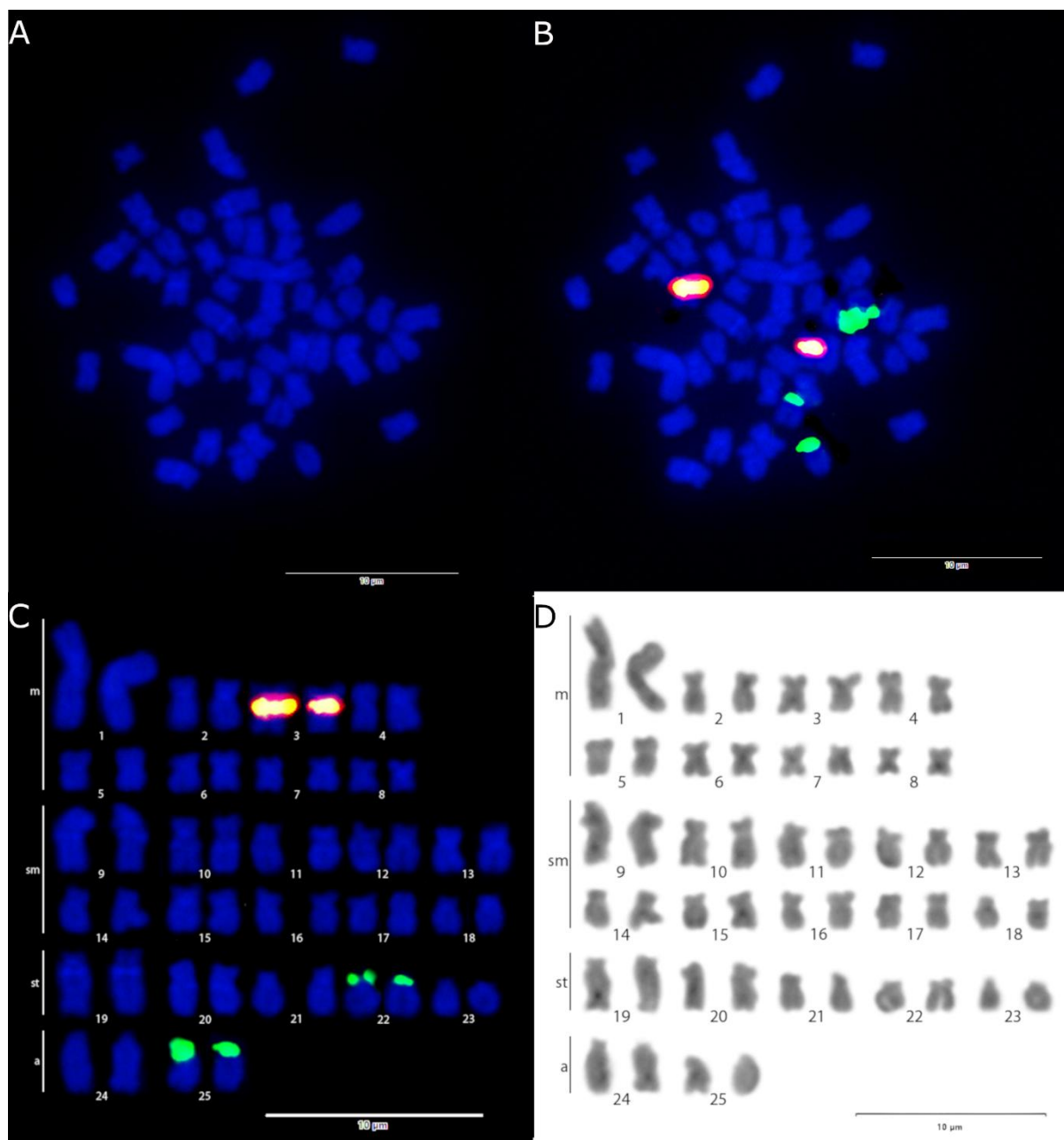


Figura 7: (a) Cariótipo de *S. Papilliferus* com marcações de genes DNAr 18S (verde) e 5S (vermelho) e (b) cariótipo da mesma metáfase para comparação.

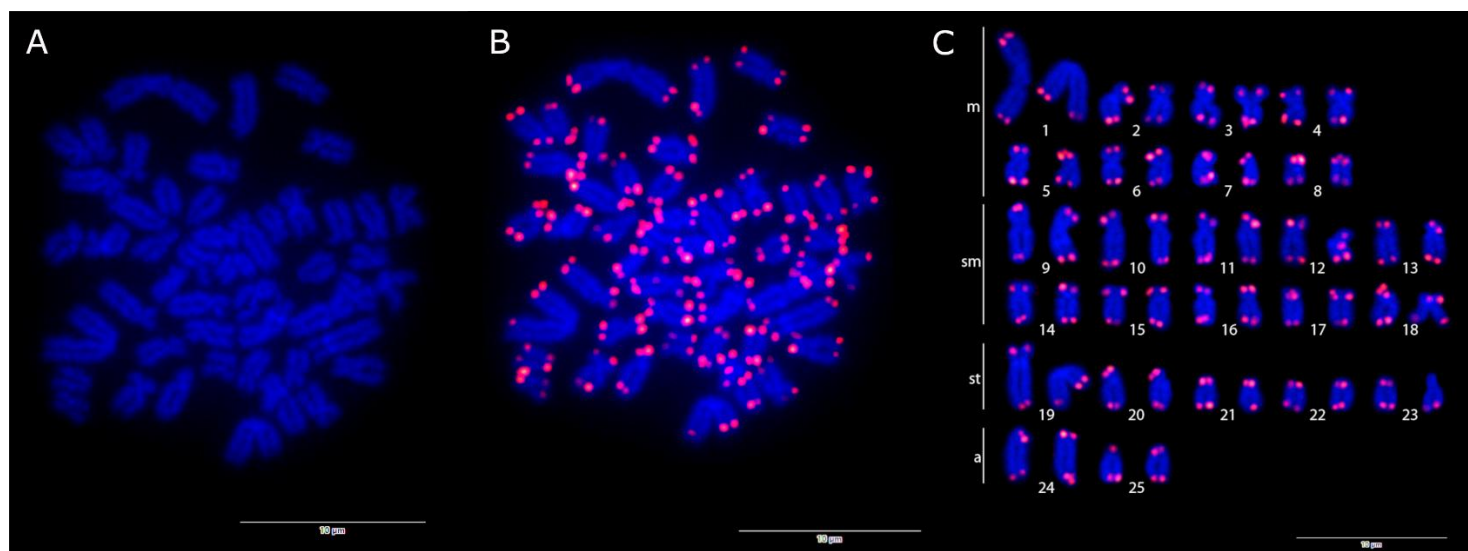


Figura 8: Cariótipo de *S. Papilliferus* com marcações do gene telomérico.

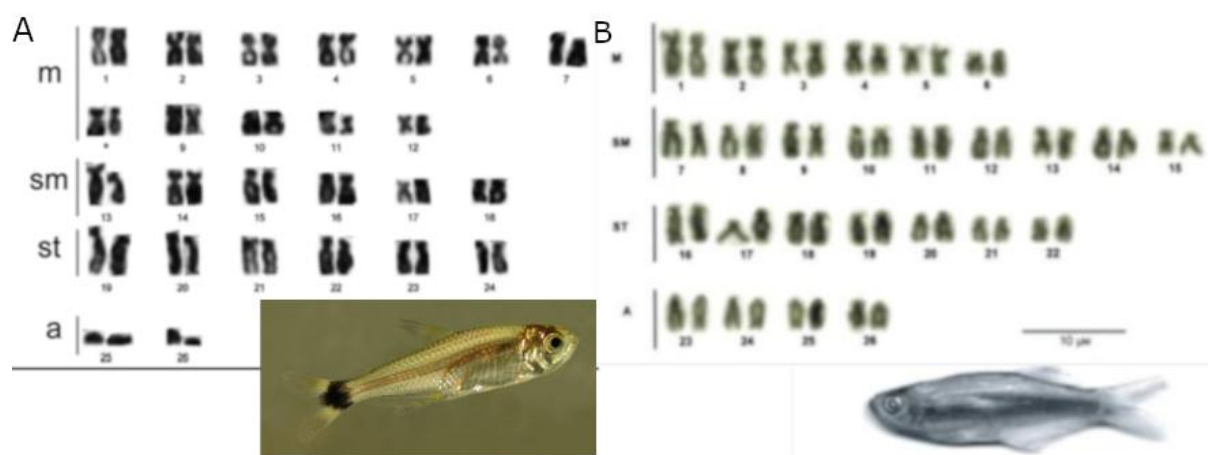


Figura 9: (A) Adaptado de Troy (2010) cariótipo de *O. pequirá*, (B) adaptado de Torres (2008) cariótipo de *M. microlepis*. Observe que ambas as espécies possuem 52 cromossomos com a ausência do primeiro grande par metacêntrico.

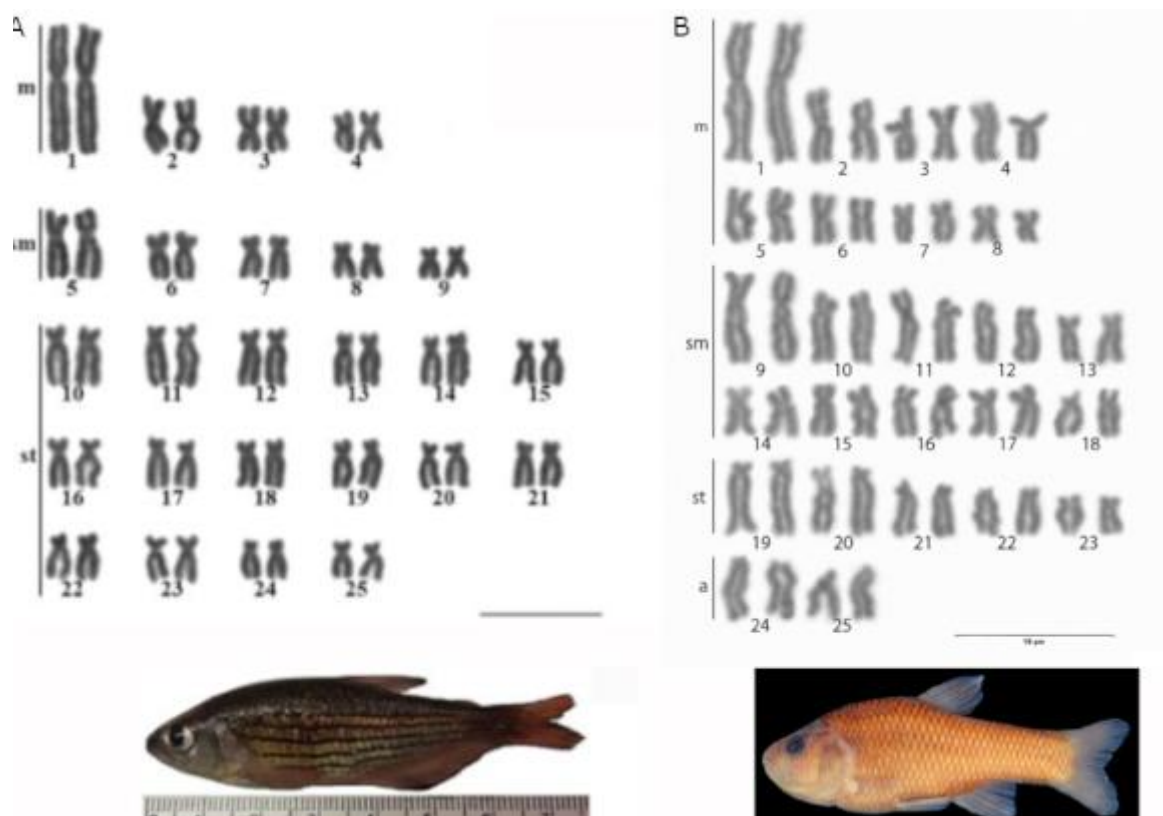


Figura 10: (A) Adaptado de Soares *et al.*, (2021) cari3tipo de *H. multifasciatus*, (B) cari3tipo de *S. papilliferus* desenvolvido neste projeto. Observe ambas as esp3cies possuem 50 cromossomos com a presen3a do primeiro grande par metac3ntrico.

5. CONCLUS3ES

O presente estudo representou um avan3o significativo no entendimento da gen3tica e de *Spintherobolus papilliferus*, uma esp3cie rara e amea3ada de extin33o, end3mica da Mata Atl3ntica. Atrav3s de uma abordagem integrada que combinou t3cnicas modernas de sequenciamento gen3mico e an3lises citogen3ticas, foi poss3vel gerar dados in3ditos e de alto valor cient3fico sobre essa esp3cie.

A montagem do draft genoma de *S. papilliferus* 3 um primeiro passo para a obten33o de genoma de refer3ncia e o primeiro genoma montado para uma esp3cie da antiga fam3lia Characidae com uma posi33o filogen3tica importante, utilizando uma plataforma de sequenciamento de leitura curta, realizada a partir de DNA gen3mico extra3do de um 3nico indiv3duo.

Esses resultados, in3ditos na literatura, fornecem *insights* importantes sobre a estrutura cromoss3mica de *S. papilliferus* e sua rela33o com outras esp3cies da antiga fam3lia Characidae. Al3m disso, a hibridiza33o *in situ* fluorescente (FISH) permitiu a localiza33o de

genes ribossomais (DNAr 18S e 5S) e a marcação de telômeros, contribuindo para uma compreensão mais detalhada da organização genômica dessa espécie.

A construção de um genoma de referência de alta qualidade, ainda em andamento, representa um marco importante para estudos futuros de genômica comparativa. Esse genoma servirá como base para investigações sobre a evolução de grandes grupos de peixes neotropicais, especialmente dentro da família Characidae, e permitirá a identificação de elementos genéticos conservados e divergentes. Além disso, a disponibilização dos dados gerados em bancos públicos, como o GenBank, e dos scripts de análise em repositórios como o GitHub, garantirá o acesso irrestrito da comunidade científica a essas informações, promovendo colaborações e avanços futuros na área.

Por fim, este estudo não apenas contribui para o conhecimento científico sobre *S. papilliferus*, mas também reforça a importância da integração de dados genômicos e citogenéticos para a conservação de espécies ameaçadas. A divulgação dos resultados em artigos científicos, congressos e simpósios permitirá que a comunidade científica e a sociedade em geral reconheçam a relevância da biodiversidade local e dos esforços para sua preservação. A continuidade dessa linha de pesquisa, com foco na genômica comparativa e na evolução cromossômica, abrirá novas perspectivas para o entendimento da diversidade genética e dos processos evolutivos que moldaram os peixes neotropicais.

6. DIFICULDADES ENCONTRADAS

A coleta de espécimes de *S. papilliferus* apresentou desafios logísticos significativos. Foram necessárias duas expedições para localizar o ponto de coleta adequado, evidenciando a complexidade do acesso ao habitat da espécie.

A obtenção de cromossomos mitóticos por meio da técnica direta revelou-se particularmente desafiadora. Apesar da aplicação de estímulos para indução da mitose, o número de metáfases obtidas foi limitado. Tal dificuldade pode ser atribuída ao porte reduzido dos indivíduos, o que dificulta a extração do tecido renal necessário para a técnica, além da ausência de protocolos padronizados para essa espécie específica, especialmente no que tange aos tempos de exposição à colchicina e à solução hipotônica.

A tentativa de cultivo celular para posterior análise citogenética não foi bem-sucedida devido a instabilidades no equipamento, que comprometeram as condições ideais para o procedimento. Adicionalmente, compromissos acadêmicos internacionais coincidiram com o período de cultivo, limitando o tempo disponível para ajustes e repetições. Diante dessas

circunstâncias, priorizou-se a obtenção de metáfases por meio da técnica direta, visando otimizar o tempo e os recursos disponíveis.

7. REFERÊNCIAS

- Allendorf FW, Hohenlohe PA, Luikart G. **Genomics and the future of conservation genetics**. Nat Rev Genet. 2010 Oct;11(10):697-709. doi: 10.1038/nrg2844. PMID: 20847747.
- Aparicio, S. *et al.* (2002). **Whole-Genome Shotgun Assembly and Analysis of the Genome of *Fugu rubripes***. Science, 297: 1301-1310.
- Arai, R. (2011). **Fish Karyotypes: A Check List**. Japan: Springer.
- Ballouz, S.; Dobin, A.; & Gillis, J. A. (2019). **Is it time to change the reference genome?**. Genome biology, 20(1), 1-9.
- Bejar J, Borrego JJ, Alvarez MC (1997) **A continuous cell line from the cultured marine fish fillet-head seabream (*Sparus aurata* L.)**. Aquaculture; 150: 143–153.
- Betancur-R, R. *et al.* (2019). **Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes**. Evolution, 73:329-345.
- Brandies, P.; Peel, E.; Hogg, C. J.; Belov, K. (2019). **The value of reference genomes in the conservation of threatened species**. Genes, 10(11), 846.
- Chaisson, M.J.P.; Wilson, R.K.; Eichler, E.E. (2015). **Genetic variation and the de novo assembly of human genomes**. Nature Reviews Genetics 16(11). DOI: 10.1038/nrg3933.
- Challis, R.; Richards, E.; Rajan, J.; Cochrane, G.; Blaxter, M. (2020). **BlobToolKit—interactive quality assessment of genome assemblies**. G3: Genes, Genomes, Genetics, 10(4), 1361-1374.
- Challis *et al.* (2023) **Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life**. Wellcome Open Res 2023, 8:24.
- Chakraborty, M.; Baldwin-Brown, J.G.; Long, A.D.; Emerson, J.J. (2016) **Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage**. Nucleic Acids Res 44:.
- Chen, Y.; Zhang, Y.; Wang, A.Y.; Gao, M.; Chong, Z. (2021). **Accurate long-read de novo assembly evaluation with Inspector**. Genome Biology, 22(1), 1-21.
- Cheng, H. *et al.* (2021) **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm**. Nat Methods 18, 170–175.
- Durand, N.C. *et al.* (2016). **Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments**. Cell Syst. Jul;3(1):95-98.

- Dudchenko, O. et al. (2017). **De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds.** *Science* 356, 92-95.
- Ekblom, R.; Wolf, J.B.W. (2014) **A field guide to whole-genome sequencing, assembly and annotation.** *Evol Appl* 7:1026–1042.
- Ellegren, H. (2014) **Genome sequencing and population genomics in non-model organisms.** *Trends Ecol Evol* 29:51–63.
- Eschmeyer, W.N.; Fong, J.D. (2023) **Catalog of fishes.** Disponível em <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>. Acessado em 22 de março de 2024.
- Felsenstein J., **Inferring Phylogenies.** Sinauer Assoc., 2004, pp. xx + 664. February 2005 *Journal of Classification* 22(1):139-142. DOI:10.1007/s00357-005-0009-4
- Foresti, F.; Almeida-tolcdo, L.F.; Toledo, S.A. (1989) **Supernumerary chromosome system, C-banding pattern characterization and multiple nucleolus organizer regions in *Moenkhausia sanctaefilomenae* (Pisces, Characidae).** *Genetica*, v. 79, n. 2, p. 107-114, 1989.
- Gelsinger, D.R. et al. (2020) **Regulatory noncoding small RNAs are diverse and abundant in an extremophilic microbial community.** *mSystems* 5, e00584-19.
- Genome 10K Community Sci. (2009) **Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species.** *J. Hered.* 100(6):659–74.
- Guiglielmoni, N. et al. (2024) **Revisiting genomes of non-model species with long reads yields new insights into their biology and evolution.** *Frontiers in Genetics*. 2024. 15 <https://doi.org/10.3389/fgene.2024.1308527>
- Guimarães, I.N. et al. (1995) **Cytogenetic studies of three species of Glandulocaudinae (Pisces, Characiformes, Characidae).** *Genetics and Molecular Biology* 18: 185–189.
- Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. (2013). **QUAST: quality assessment tool for genome assemblies.** *Bioinformatics*, 29(8), 1072-1075.
- Howell, W. M., Black, D. A. (1980) **Controlled silver staining of nucleolus organizer regions with a protective colloidal developer: a 1-step method.** *Experientia* 36, 1014–1015.
- Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio). (2022) ***Spintherobolus papilliferus*.** *The IUCN Red List of Threatened Species* 2022: e.T187287A1825924.
- Jain, M. et al. (2018) **Nanopore sequencing and assembly of a human genome with ultra-long reads.** *Nat Biotechnol* 36:338–345.

- Koepfli, K.P. *et al.* (2015) **The Genome 10K Project: A Way Forward**. *Annu. Rev. Anim. Biosci.* 3:57–111. Doi: 10.1146/annurev-animal-090414-014900.
- Lee, Y.-C. *et al.* (2023) **Single-worm long-read sequencing reveals genome diversity in free-living nematodes**. *Nucleic Acids Research.* 51(15):8035–8047. <https://doi.org/10.1093/nar/gkad647>.
- Li, H. (2018) **Minimap2: pairwise alignment for nucleotide sequences**. *Bioinformatics.* Sep 15;34(18):3094-3100.
- Li, H.; Durbin, R. (2009) **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics.* Jul 15;25(14):1754-60.
- Lieberman-Aiden, E. *et al.* (2009) **Comprehensive mapping of long-range interactions reveals folding principles of the human genome**. *Science* 326: 289–293.
- Mattox G.M.T. *et al.* **Spintherobolus papilliferus in the rio Ribeira de Iguape basin: implications for the biogeography and conservation of an endangered species (Ostariophysi: Characiformes)**. *Neotrop Ichthyol.* 2023; 21(2):e230026.
- Melo, B.F. *et al.* (2021). **Accelerated Diversification Explains the Exceptional Species Richness of Tropical Characoid Fishes**. *SYSTEMATIC BIOLOGY*, v. 71, p. 78-92.
- Mirande, J.M. (2019). **Morpholgy, molecules and the phylogeny of Characidae (Teleostei: Characiformes)**. *Cladistics*, 35(3): 282-300.
- Netto, M.R.C.B. *et al.* (2007) **A standard protocol for obtaining fish chromosomes under post-mortem conditions**. *Micron* 38: 214-217.
- Nielsen R, *et al.* (2005) **Genomic scans for selective sweeps using SNP data**. *Genome Res.* Nov;15(11):1566-75. doi: 10.1101/gr.4252305. PMID: 16251466; PMCID: PMC1310644.
- Nurk, S. *et al.* (2022). **The complete sequence of human genome**. *Science*, 376 (6588): 44-53.
- Oliveira, C. *et al.* (2011) **Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes): a multilocus phylogenetic approach**. *BMC Evolutionary Biology* 2011, 11:275.
- Okonechnikov, K., Conesa, A., García-Alcalde, F. (2016). **Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data**. *Bioinformatics*, 32(2), 292-294.
- Paim, F.G.; Maia, L.; Landim-Alvarenga, F.C.; Foresti, F.; Oliveira, C. (2018) **New Protocol for Cell Culture to Obtain Mitotic Chromosomes in Fishes**. *Methods Protoc.* 1, 47.
- Phillippy, A.M. (2017) **New advances in sequence assembly**. *Genome Res.* 2017 27: xi-xiii.

- Pinkel, D., Straume, T., and Gray, J. W. (1986). **Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization.** Proc. Natl. Acad. Sci.U.S.A. 83, 2934–2938
- Rhie, A. *et al.* (2021). **Towards complete and error-free genome assemblies of all vertebrate species.** Nature, 592(7856), 737-746.
- Rhoads, A.; Au, K.F. (2015) **PacBio Sequencing and Its Applications.** Genomics Proteomics Bioinformatics. Oct;13(5):278-89. doi: 10.1016/j.gpb.2015.08.002. Epub 2015 Nov 2. PMID: 26542840; PMCID: PMC4678779.
- Romanenko SA, Biltueva LS, Serdyukova NA, Kulemzina AI *et al.* (2015) **Segmental paleotetraploidy revealed in starlet (*Acipenser ruthenus*) genome by chromosome painting.** Molecular Cytogenetics, 8:90.
- Ruiz-Herrera, A.; Farré-Belmonte, M. (2021) **Mechanisms Driving Karyotype Evolution and Genomic Architecture.** Mech Driv Karyotype Evol Genomic Archit.
- Sánchez-Romero, O. *et al.* (2015) **First description of the karyotype and localization of major and minor ribosomal genes in *Rhoadsia altipinna* Fowler, 1911 (Characiformes, Characidae) from Ecuador.** Comparative Cytogenetics 9(2): 271–280.
- Sedlazeck, F.J.; Lee, H.; Darby, C.A.; Schatz, M.C. (2018). **Piercing the dark matter: bioinformatics of long-range sequencing and mapping.** Nat Rev Genet. 19(6):329-346.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. (2015). **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** Bioinformatics, 31(19), 3210-3212.
- Shifu Chen. (2023) **Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp.** iMeta 2: e107. <https://doi.org/10.1002/imt2.107>
- Silva, D.M.Z.A. *et al.* (2021). **Long-term persistence of supernumerary B chromosomes in multiple species of *Astyanax* fish.** BMC Biology (2021) 19:52
- Soares, L.B *et al.* (2021) **Molecular cytogenetic analysis and the establishment of a cell culture in the fish species *Hollandichthys multifasciatus* (Eigenmann & Norris, 1900) (Characiformes, Characidae).** Animal Genetics • Genet. Mol. Biol. 44 (2). <https://doi.org/10.1590/1678-4685-GMB-2020-0260>
- Souza, I.L.; Moreira-Filho, O.; Bertollo, L.A.C. (1995) **Karyotypic characterization of *Aphyocharax difficilis* (Pisces, Characidae): C-banding, Ag-NORs and occurrence of diplochromosomes.** Cytobios 83: 33–39.
- Sumner AT (1972) **A simple technique for demonstrating centromeric heterochromatin.** Exp. Cell Res. 75, 304–306.

- Torres, R.A. *et al.* (2008) **Chromosomes, RAPDs and evolutionary trends of the Neotropical fish *Mimagoniates microlepis* (Teleostei: Characidae: Glandulocaudinae) from coastal and continental regions of the Atlantic forest, Southern Brazil.** Acta Zoologica, 89: 253-259. <https://doi.org/10.1111/j.1463-6395.2007.00314.x>
- Troy, W.P. *et al.* (2010) **Caracterização do número cromossômico em espécies de *Odontostilbe* e *Serrapinnus* (characidae: cheirodontinae) da Bacia do rio Paraguai, Brasil.** Publ. UEPG Ci. Biol. Saúde, Ponta Grossa, v.16, n.1, p. 57-61
- Ureta-Vidal A, Ettwiller L, Birney E. (2003) **Comparative genomics: genome-wide analysis in metazoan eukaryotes.** Nat Rev Genet. Apr;4(4):251-62. doi: 10.1038/nrg1043. PMID: 12671656.
- Utsonomia, R. *et al.* (2019). **Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome.** Scientific Reports, 9:5856.
- Vertebrate Genomes Project. <https://vertebrategenomesproject.org/technology> Acesso em 02 de junho de 2023.
- Vignieri, S. (2023). **Understanding our own order.** Science, 380 (6648): 904-905.
- Wang X, Yang J, Chen X, Pan X (2012) **Establishment and characterization of a fibroblast-like cell line from *Anabarrilius grahami* (Cypriniformes: Cyprinidae).** Zoological Research 33, (E5-6): E89–97.
- Warren, W.C. *et al.* (2021). **A chromosome-level genome of *Astyanax mexicanus* surface fish for comparing population-specific genetic differences contributing to trait evolution.** Nature Communications, 12:1447.
- Waterhouse, R. M. *et al.* (2018). **BUSCO applications from quality assessments to gene prediction and phylogenomics.** Molecular biology and evolution, 35(3), 543-548.
- Zhang Q, Cooper RK, Wolters WR, Tiersch TR (1998) **Isolation, culture and characterization of a primary fibroblast cell line from channel catfish.** Cytotechnology, 26(2): 83–90.