

GUSTAVO DE ALMEIDA CARON

**Caracterização do gene snDNA U4 em peixes da ordem Characiformes
(Vertebrata: Teleostei)**

Bauru 2024

GUSTAVO DE ALMEIDA CARON

**Caracterização do gene snDNA U4 em peixes da ordem Characiformes
(Vertebrata: Teleostei)**

Dissertação apresentada à Universidade Estadual Paulista UNESP, Faculdade de Ciências, Bauru, para a obtenção do título de Mestre em Biociências.

Área de Concentração: Caracterização da diversidade biológica.

Orientador: Prof. Dr. Ricardo Utsunomia.

Bauru

2024

C293c

Caron, Gustavo de Almeida

Caracterização do gene snDNA U4 em peixes da ordem
Characiformes (Vertebrata: Teleostei) / Gustavo de Almeida Caron. --
Bauru, 2024

38 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP),
Faculdade de Ciências, Bauru

Orientador: Ricardo Utsunomia

1. Genes U. 2. Genes. 3. Bioinformática. 4. Peixes neotropicais. 5.
Peixes de água doce. I. Título.

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

Essa ficha não pode ser modificada.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE GUSTAVO DE ALMEIDA CARON, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCIÊNCIAS, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.

Aos 10 dias do mês de maio do ano de 2024, às 09:30 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de GUSTAVO DE ALMEIDA CARON, intitulada **Caracterização do gene snDNA U4 em peixes da ordem Characiformes (Vertebrata: Teleostei)**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. RICARDO UTSUNOMIA (Orientador(a) - Participação Virtual) do(a) Departamento de Ciências Biológicas / Faculdade de Ciências - Unesp/Câmpus de Bauru, Prof. Dr. FABIO PORTO FORESTI (Participação Virtual) do(a) Departamento de Ciências Biológicas / Faculdade de Ciências - Unesp/Câmpus de Bauru, Prof. Dr. ROBERTO FERREIRA ARTONI (Participação Virtual) do(a) Departamento de Biologia Estrutural Molecular e Genética / Universidade Estadual de Ponta Grossa. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final: Aprovado . Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Prof. Dr. RICARDO UTSUNOMIA

Documento assinado digitalmente
 **RICARDO UTSUNOMIA**
Data: 15/05/2024 11:02:58-0300
Verifique em <https://validar.itl.gov.br>

Agradecimentos

Primeiramente, quero agradecer aos meus pais, Sueli e Caron, por me darem o presente da vida e sustentarem meu desenvolvimento, me dando abrigo, nutrição, afeto, suporte, e exemplo, permitindo a mim ser o que sou hoje.

Quero agradecer a Natureza e sua exuberante teia de relações, sua complexidade e beleza impressionantes. Ao interesse que tive por ela desde menino, observando as formas de vida no quintal de minha avó, quintal que me expôs a esse contato a natureza, sem o qual talvez jamais tivesse desenvolvido esse apreço, me levando a querer estudá-la e sendo minha maior fonte de inspiração e de sentido existencial.

Quero agradecer a todos os meus amigos, familiares e à minha companheira, que de alguma forma me deram forças e incentivo para a execução deste projeto.

Agradeço muito a ajuda do pessoal do laboratório, sem os quais eu não teria conseguido realizar este trabalho.

Agradeço a também à Universidade Estadual Paulista Júlio de Mesquita Filho – UNESP, por existir e ser o palco deste e tantos outros projetos relevantes.

Por último, agradeço meu orientador Prof. Dr. Ricardo Utsunomia, por me conceder esta oportunidade a qual sonhei por tanto tempo, tornando possível a realização deste projeto, por todo ensinamento e a oportunidade de conviver com profissionais tão qualificados, pelo olhar humano, a compreensão e paciência que me permitiram realizar este projeto mesmo dentro da realidade de uma exigente rotina de trabalho. Muito obrigado.

CARON, Gustavo de Almeida. **Caracterização do gene snDNA U4 em peixes da ordem Characiformes (Vertebrata: Teleostei)**. 2024. 36f. Dissertação (Mestrado Acadêmico em Biociências), Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Assis, 2024.

Resumo

A enorme diversidade da biosfera terrestre se deve, em grande parte, à evolução do *splicing*, mecanismo molecular presente em todos os eucariotos que é responsável pela retirada dos íntrons para a produção do RNAm maduro. No centro da maquinaria molecular responsável por este processo, estão os pequenos RNAs nucleares (snRNAs), transcritos pelos genes Us, como os snDNAs U1, U2, U4, U5 e U6, que estão organizados em tandem nos genomas de eucariotos, onde cópias gênicas e espaçadoras são repetidas dezenas de vezes de maneira enfileirada. Até meados da década passada, a caracterização da organização de genes arranjados desta forma era bastante laboriosa, exigindo uma rotina de tecnologia do DNA recombinante para clonar e sequenciar unidades parciais destas sequências. Com o progresso e popularização de técnicas de sequenciamento massivo, em conjunto com o desenvolvimento de ferramentas bioinformáticas adequadas, a caracterização deste tipo de sequência em espécies não-modelo se tornou viável. Nesse sentido, o objetivo deste trabalho foi caracterizar as unidades repetidas em tandem do snDNA U4 em diferentes espécies de peixes Characiformes a partir de uma *pipeline* semiautomatizada, utilizando *reads* curtos brutos (2x100nt). A aplicação deste método permitiu a caracterização completa do snDNA U4 (região codificadora + espaçadora) para todas as espécies escolhidas. Os resultados revelaram que a região codificadora é bastante conservada entre espécies, enquanto a região espaçadora é muito variável, tanto em tamanho (de 664nt a 1464nt) quanto em composição. Do ponto de vista de variação intragenômica, é possível notar que mais de 90% dos polimorfismos estão localizados nas regiões espaçadoras, conforme era esperado pela teoria, uma vez que alterações nas regiões de transcrição geralmente são eliminadas pelo processo de seleção purificadora. Os resultados obtidos aqui indicam que *pipelines* apropriadas são excelentes alternativas à métodos de biologia molecular clássica, como clonagem e sequenciamento de Sanger.

Caron, Gustavo de Almeida. **Characterization of the U4 snDNA gene in fish of the order Characiformes (Vertebrata: Teleostei)**. 2024. 36p. Dissertation (Masters in Bioscience) – São Paulo State University (UNESP), School of Science, Humanities, and Languages, Assis, 2024.

Abstract

The vast diversity of the terrestrial biosphere is largely attributed to the evolution of splicing, a molecular mechanism present in all eukaryotes responsible for removing introns to produce mature mRNA. At the core of the molecular machinery governing this process are small nuclear RNAs (snRNAs), transcribed by U genes, such as U1, U2, U4, U5, and U6, organized in tandem in eukaryotic genomes, where gene copies and spacers are repeated dozens of times in a linear arrangement. Until the mid-2000s, characterizing the organization of genes arranged in this manner was laborious, requiring a routine of recombinant DNA technology to clone and sequence partial units of these sequences. With the progress and popularization of massive sequencing techniques, coupled with the development of suitable bioinformatic tools, characterizing this type of sequence in non-model species has become feasible. In this regard, the objective of this study was to characterize the tandemly repeated units of U4 snDNA in different Characiformes fish species using a semi-automated pipeline, utilizing raw short reads (2x100nt). The application of this method allowed for the complete characterization of the U4 snDNA (coding region + spacer) for all chosen species. Results revealed that the coding region is highly conserved across species, while the spacer region is highly variable, both in size (ranging from 664nt to 1464nt) and composition. From an intragenomic variation perspective, it is notable that over 90% of polymorphisms are located in the spacer regions, as expected by theory, since alterations in transcription regions are typically eliminated by purifying selection. The obtained results indicate that appropriate pipelines are excellent alternatives to classical molecular biology methods, such as cloning and Sanger sequencing.

Sumário

1. Introdução	
1.1 Famílias multigênicas, características e organização genômica.....	9
1.2 Genes snDNA U.....	12
1.3 A bioinformática como ferramenta para estudos em citogenômica.....	13
1.4 Peixes neotropicais e diversidade genética.....	13
2. Objetivos.....	15
2.1 Objetivos gerais.....	15
2.2 Objetivos específicos.....	15
3. Material e métodos.....	16
3.1 Seleção das bibliotecas genômicas e análises bioinformáticas.....	16
3.2 Identificação do gene snRNA U4 e alinhamento das sequências semelhantes....	17
3.3 Identificação de SNPs (<i>Single Nucleotide Polimorphisms</i>).....	18
3.4 Avaliação de abundância das unidades de snDNA U4.....	18
3.5 Análises comparativas com genomas completos.....	19
4. Resultados.....	20
4.1 Montagem de unidades de snDNA U4.....	20
4.2 Identificação e análise de SNPs.....	22
4.3 Análise de cobertura.....	23
4.4 Análise comparativa com genomas completas.....	25
5. Discussão.....	26
6. Referências.....	27

1. Introdução

1.1. Famílias multigênicas, características e organização genômica

Os genomas de organismos eucariotos são compostos por diferentes classes de DNA, sendo, boa parte formados por sequências de cópia única ou sequências repetitivas, que se repetem desde algumas dezenas até milhares de vezes (Sumner 2003; López-Flores e Garrido-Ramos 2012). As sequências repetitivas podem se apresentar dispersas pelo genoma, como os elementos transponíveis (Yuan 2018), ou organizadas em tandem, como os DNAs satélites e famílias multigênicas (Charlesworth et al. 1994; Garrido-Ramos 2017). Na maioria das espécies eucariotas, as sequências repetitivas representam mais da metade de todo conteúdo genômico (Biscotti et al. 2015), além de desempenharem um papel importante em variações cariotípicas em muitos organismos (Kidwell 2002, Biscotti et al. 2015). Dentre estas sequências, há aquelas que são não codificantes, como os DNAs satélites (satDNAs), e as codificantes, como as famílias multigênicas; a exemplo destas, temos os genes ribossômicos, os genes de histonas, e os genes snDNA Us (Lopes-Flores e Garrido-Ramos 2015).

As famílias multigênicas apresentam-se dispostas em um número relativamente alto de cópias dispostas em tandem, em um ou mais clusters por genoma, características que as tornam bastante adequadas para técnicas de mapeamento cromossômico, fato corroborado pelos inúmeros estudos em que foram utilizadas como marcadores cromossômicos para iluminar questões como evolução cariotípica e organização genômica (Nguyen et al. 2010; Cabral-de-Mello et al. 2011; Garcia-Souto et al. 2018; Mazzoleni et al. 2018; Anjos et al. 2019; Degrandi et al. 2020).

Os satDNAs diferem das famílias multigênicas por apresentarem um único tipo de sequência, enquanto as famílias multigênicas são subdivididas em porções com funções distintas, como genes e espaçadores (Ruiz-Ruano et al. 2016). Outra distinção é que as famílias multigênicas apresentam genes similares em estrutura e função, enquanto os satDNAs apresentam sequências não codificantes e ainda pouco, ou nada, compreendidas em suas funções (Nei e Rooney 2005). Supõe-se que as famílias multigênicas originam-se a partir de um gene ancestral que se duplica, dando início a múltiplas cópias de genes parálogos no genoma do organismo (Innan e Kondrashov 2010), sendo essa duplicação compreendida como o processo central na evolução de famílias multigênicas, uma vez que as células necessitam de quantidades significativas de seus produtos gênicos (Sumner 2003), além de outros processos envolvidos na

evolução desta classe, como a pseudogenização, recombinação e seleção natural (Eirín-Lopez et al. 2012).

Basicamente, as famílias multigênicas estão envolvidas na codificação de importantes RNAs e proteínas que atuam no metabolismo e organização celular. Juntamente com outras categorias de DNAs repetitivos, estes elementos também estão envolvidos na organização e funcionalidade dos centrômeros, telômeros, segregação cromossômica, regulação gênica, reparo e replicação do DNA, além de trabalhos relatando sua participação na diferenciação dos cromossomos sexuais (Anteleitner e Haymer 1992; Kraemer e Schmid 1993; Messias et al. 1996, Martins et al. 2007).

A família multigênica dos genes U, ou snDNA (do inglês, *small nuclear DNA*), está envolvida no processo de *splicing*. Fazem parte desta família os genes do complexo spliceossomal principal, sendo U1, U2, U4, U5 e U6 (Bringmann e Luhrmann, 1986; Nilsen, 2003; Valadkhan, 2005; West, 2012). A respeito de sua localização no genoma, ao contrário dos genes ribossômicos DNAr 45S e 5S, que se mostram bastante dinâmicos, com variações cromossômicas entre espécies/populações (Cabrero e Camacho 2008; Cabrero et al. 2009; Cioffi et al. 2010; Cabral-deMello et al. 2011; Rebordinos et al. 2013), os genes snDNAs se mostram mais conservados em número e localização cromossômica (Nei e Rooney 2005; Cabral-de-Mello et al. 2012; Utsunomia et al. 2014; Anjos et al. 2015; Silva et al. 2015), embora deva-se ter em mente que foram feitos muito mais estudos envolvendo os genes de DNAr do que de snDNAs.

Como exemplos bem conhecidos de famílias multigênicas temos os genes codificadores de histonas (H1, H2A, H2B, H3 e H4), organizados em um cluster gênico espaçados por DNAs não codificantes, podendo, em outras espécies, também apresentar-se de maneira isolada (Lifton et al. 1978; Engel e Dogson 1981; Ruberti et al. 1982). A família multigênica das histonas codificam pequenas proteínas que constituem a maior parte da cromatina e têm como principal função o auxílio no empacotamento do DNA (van Holde 1988; Jenuwein and Allis 2001).

Os RNAs ribossomais constituem outra família multigênica composta por unidades repetitivas dispostas em tandem, como o gene para RNAr 5S, espaçado por sequências não transcritas extremamente variadas e o gene de RNAr 45S, que é transcrito nos RNAs ribossomais 18S, 5.8S e 28S, sendo estas sequências separadas por espaçadores intergênicos transcritos (ITS *Internal Transcribed Spacer*) com os cluster separados por espaçadores intergênicos e por espaçadores transcritos externos (ETS-*External Transcribed Spacer*) (Long e

Dawid 1980; Williams e Strobeck 1985; Martins e Wasko 2004; Eickbush e Eickbush 2007), sendo muitas destas cópias correspondentes à RONS (Regiões Organizadoras de Nucléolos), as quais em peixes têm demonstrado grande variabilidade (Foresti et al. 1981; Almeida-Toledo 1998). Apesar de seu papel central no metabolismo celular e relativa conservação de suas sequências, foram observadas altas taxas de diversificação molecular e cromossômica em curtas escalas de tempo nos genes de DNAr em algumas espécies (Cabrero et al. 2003; Datson e Murray 2006; Ferretti et al. 2019). O DNAr 5S é geralmente encontrado em clusters mais ou menos similares de arranjos repetidos em tandem, com as principais diferenças ocorrendo nos espaçadores não transcritos (NTS) (Shippen- Lentz e Vezza 1988).

Vierna et al. (2011) analisaram as famílias de DNAr 5S e snDNA U1 em 10 espécies de 4 gêneros de moluscos bivalves da família Pharidae, obtendo a primeira sequência de U1 em bivalves, e revelaram o *linkage* entre ambas as famílias multigênicas. O DNAr 5S é frequentemente encontrado unido a outras famílias multigênicas, tendo este padrão sido detectado praticamente em todos os grupos eucariotos (Cross e Rebordinos, 2005; Freire et al. 2010; Cabral-de-Mello et al. 2010), enquanto o *linkage* entre o DNAr 5S e o snDNA U1 tem sido relatado em diferentes grupos (Pelliccia et al. 2001; Manchado et al. 2006; Yano et al. 2020). Essa falta de conservação e a diversidade dos tipos de ligações apontam uma improbabilidade desses acontecimentos terem proporcionado algum tipo de seleção vantajosa (Drouin e Moniz de Sá 1995). Marz et al. (2008) concluíram que sequências repetidas em tandem de diferentes famílias de snDNA, ou snDNA unidos a DNAr 5S, não são conservadas por longos períodos evolutivos em metazoários.

Devido ao processo de evolução em concerto ser uma das formas de evolução atribuídas as famílias multigênicas (Eickbush e Eickbush 2007), acreditava-se que o DNAr 5S fosse caracterizado por baixos níveis de divergência intragenômica em praticamente todas as espécies. No entanto, variantes de DNAr 5S em genomas de animais, plantas e fungos foram identificadas (Fernandez et al. 2005; Rooney e Ward 2005; Caradonna et al. 2007), e, ainda, alguns casos em que se observou diferenças nas regiões codificadoras de RNAs, correspondendo a variantes tecidoespecíficas (Peterson et al. 1980). Desta forma, concebeu-se que a evolução de famílias multigênicas se dá por uma mistura de processos, como a evolução em concerto e *birth-and-death* (Nei and Rooney, 2005; Mount et al. 2007; Freire et al. 2010; Vierna et al. 2010; Pinhal et al. 2011; Merlo et al. 2012; Bardella e Cabral-de-Mello 2018; Zhang et al. 2021).

A evolução em concerto consiste em um processo no qual membros da mesma família gênica evoluem de maneira conjunta, resultando na homogeneização das unidades (Dover 1982; Ugarkovic e Plohl 2002; Eickbush e Eickbush 2007; Plohl et al. 2008; Garrido-Ramos 2017), sendo os processos de conversão gênica e *crossing-over* desigual, os principais mecanismos que atuam nesse processo (Smith 1976; Walsh 1987; Dover 2002; Shi et al. 2010). Por outro lado, a evolução *birth-and-death* consiste em um processo no qual novas sequências são geradas por duplicação gênica e escapam do processo de homogeneização, podendo persistir no genoma por longos períodos ou se perderem via *crossing-over* desigual, com as cópias que persistirem podendo então sofrer divergência por mutações ou pseudogenização (Hughes e Nei 1992; Nei e Rooney 2005; Eirin-Lopez et al. 2012).

1.2 Genes snDNA U

O mecanismo de *splicing*, é uma etapa fundamental da expressão gênica em eucariotos. Grande parte dos genes transcritos pela RNA polimerase II é expressa como mRNA precursores (pré-RNAs), que são então processados em mRNAs maduros pelo mecanismo de *splicing*, que retira regiões não codificantes (íntrons) e conecta as regiões codificantes (éxons) (Will e Lührmann 2011), processo realizado pela complexa maquinaria do spliceossomo, formado por cinco subunidades diferentes de ribonucleoproteínas (RNP) associadas a cofatores proteicos (Lerner et al. 1980; Will e Lührmann 2011; Shi 2017). Existem dois tipos de spliceossomos em eucariotos, cada um composto por cinco snRNAs (*small nuclear RNAs*), codificados por sequências snDNAs. O spliceossomo principal, composto pelos snRNAs U1, U2, U4, U5 e U6, que remove íntrons do tipo U2, e o spliceossomo menor (U11, U12, U4atac, U5 e Uatac), que é menos abundante e remove alguns raros íntrons dependentes de U12 (geralmente representam menos de 0,5% de todos os íntrons nos genomas) (Patel e Steitz 2003; Turunen et al. 2013); estes últimos diferem do tipo U2 pelas sequências consensos reconhecidas pela maquinaria do spliceossomo: enquanto os íntrons do tipo U2 têm terminais GT-AG (raramente AT-AC e GC-AG), o tipo U12 tem terminais AT-AC (raramente GT -AG) (Patel e Steitz 2003; Sheth et al. 2006). No entanto, esta configuração não é unânime para todos os eucariotos: alguns grupos como leveduras e nematoides podem não apresentar os componentes do spliceossomo menor, e, em outros casos, pode haver ausência de alguns genes snRNA do spliceossomo menor (Mewes et al. 1997; Burge et al. 1998; Russell et al. 2006; López et al. 2008; Marz et al. 2008).

1.3 A bioinformática como ferramenta para estudos em citogenômica

O surgimento do sequenciamento de nova geração (NGS) e o desenvolvimento de novas ferramentas de bioinformática, permitiram a caracterização *de novo* e análises de diversidade e abundância de diversos tipos de sequências de DNA repetitivo, como DNAs satélites, famílias multigênicas, e elementos transponíveis (Novak et al. 2014; Wang et al. 2015; Ruiz-Ruano et al. 2016). Deste modo, a combinação de análises físicas e *in silico* acelerou a caracterização de novas regiões genômicas e permitiu *insights* sobre a evolução e organização de famílias repetitivas de DNA (Ruiz-Ruano et al. 2016). Alguns exemplos de pipelines criadas nesse sentido, como RepeatExplorer (Novák et al. 2013), dnaPipeTE (Goubert et al. 2015), bem como o TAREAN (Novák et al. 2017), demonstram um grande desenvolvimento de ferramentas de bioinformática nessa área, fundamentais para a continuidade e aperfeiçoamento sobre os estudos envolvendo essa temática envolvendo sequências repetitivas em genomas eucarióticos. Destaca-se que a elevada abundância destas sequências permite que baixas coberturas de sequenciamento (cerca de 1x de cobertura) sejam suficientes para a caracterização da porção repetitiva de genomas de espécies não-modelo.

1.4 Peixes Neotropicais e diversidade genética

A maior diversidade de peixes no planeta, com aproximadamente 9.100 espécies identificadas, sendo cerca de 5.160 espécies presentes em águas continentais, se encontra na região Neotropical, que se estende do México ao sul da Argentina (Reis et al. 2016). Os peixes Neotropicais compõem a ictiofauna de água doce mais rica e diversificada do mundo, com as principais ordens sendo Characiformes, Siluriformes, Gymnotiformes, Cyprinodontiformes e Perciformes (Lundberg et al. 2000), sendo a ordem dos Characiformes uma das mais diversas, com 2342 espécies descritas divididas em 24 famílias, segundo Eschmeyes e Fricke (2023), com representantes dessa ordem presentes também em outras regiões como África subsaariana e América do Norte (Moreira 2007). Esse grupo reúne características citogenéticas diversas, como grande diversidade cariotípica, presença de cromossomos B e diversos sistemas sexuais, tornando um excelente grupo para estudos citogenéticos (Oliveira et al. 2009).

Toda essa diversidade da ictiofauna, expressa também geneticamente, vem sendo bem estudada e importantes informações vem sendo obtidas no campo da citogenética e citogenética molecular, com estudos promissores envolvendo evolução, identidade cariotípica, sistemas

sexuais, padrão diferencial de heterocromatina, cromossomos supranumerários, localização *in situ* de sequências de DNA ribossômico, DNAs satélites e regiões organizadoras de nucléolos (RONs), identificação de sequências específicas como sítios de DRNAr 18S e 5S, para citar alguns exemplos, sendo algumas sequências utilizadas para identificar rearranjos cromossômicos, cromossomos sexuais e caracterizar híbridos interespecíficos (Almeida-Toledo 1998; Bertollo et al. 2000; Artoni et al. 2001; Vicente et al. 2003; Vicari et al. 2008). Ainda, várias famílias de DNA repetitivo foram mapeadas em Characiformes, como famílias multigênicas como o DNA ribossomal, histonas, e mais recentemente genes U (Mestriner et al. 2000; Hashimoto et al. 2011; Silva et al. 2013, 2015, 2017; Piscor et al. 2018). No entanto, as informações sobre os genes U estão restritas a localização cromossômica das sequências de U1 e U2 (Silva et al. 2015; Piscor et al. 2018).

Dados de mapeamento envolvendo os snRNAs U1 e U2 em peixes indicam a ocorrência de apenas um ou alguns *loci* cromossômicos (Supiwong et al. 2013; Utsunomia et al. 2014; GarcíaSouto et al. 2015; Scacchetti et al. 2015; Silva et al. 2015; Piscor et al. 2018; Ponzio et al. 2018). Além disso, os snDNAs U4 e U6 foram mapeados apenas no peixe *Hollandichthys multifasciatus* (Soares et al. 2021), também com um *locus* cromossômico apenas.

Os genes snDNAs U compõem parte fundamental para o funcionamento dos organismos eucariotos, mas ainda existe uma grande carência de informações sobre a evolução e diversidade desses genes, se fazendo necessários estudos envolvendo essa temática para obtermos maiores informações sobre a distribuição, número de cópias e organização genômica dos snDNAs Us em peixes. Este trabalho visou a caracterização e comparação de unidades repetidas em tandem do gene snDNA U4 entre diferentes espécies de peixes Characiformes. Para isso, desenvolvemos um protocolo de caracterização de unidades repetidas em tandem utilizando ferramentas de bioinformática.

2. Objetivos

O presente projeto compõe o programa geral de estudos genéticos de peixes Neotropicais desenvolvidos no Laboratório de Genômica e Conservação de Peixes (Faculdade de Ciências, UNESP/Bauru), visando o conhecimento da estrutura, organização e evolução do genoma de peixes. Deste modo, os objetivos deste trabalho foram:

2.1 Objetivos gerais

- Desenvolver uma *pipeline* específica para caracterização de famílias multigênicas repetidas em tandem, utilizando *short-reads*;
- Caracterizar e comparar o gene snDNA U4 em 11 espécies de peixes da ordem Characiformes, utilizando a pipeline desenvolvida;

2.2 Objetivos específicos

Identificar polimorfismos presentes no gene snDNA U4 caracterizado em todas as espécies estudadas, bem como uma análise de cobertura para inferência da distribuição desse gene pelos genomas e possíveis unidades orfãs, e promover uma análise sobre a organização e evolução dessa família multigênica em peixes neotropicais.

3. Material e Métodos

3.1 - Seleção das bibliotecas genômicas e análises bioinformáticas em servidor remoto

Foram utilizadas bibliotecas de sequenciamento massivo de segunda geração (leituras curtas e pareadas 2x100nt) de 11 espécies de peixes da ordem Characiformes (Tabela 1). Grande parte das análises bioinformáticas e processamento de dados massivos foram realizadas em um servidor remoto localizado no Departamento de Biologia Estrutura e Funcional do Instituto de Biociências de Botucatu, sob a responsabilidade do prof. Dr. Claudio de Oliveira.

Tabela 1: Espécies cujos *clusters* de genes snRNA U4 foram caracterizados e analisados, suas respectivas famílias e número de acesso à biblioteca genômica correspondente.

Espécie	Família	Tamanho do Genoma	Biblioteca – nº de acesso
<i>Characidium gomesi</i>	Crenuchidae	1.1Gb	SRR11679749
<i>Characidium zebra</i>	Crenuchidae	1.1Gb	Biblioteca própria
<i>Colossoma macropomum</i>	Serrasalmidae	1.5Gb	SRR22219024
<i>Cyphocara modestus</i>	Curimatidae	1.6Gb	SRR26178311
<i>Hoplias malabaricus</i>	Erythrinidae	1,1Gb	SRR11676684
<i>Leporinus friderici</i>	Anostomidae	1.3Gb	SRR11676679
<i>Megaleporinus macrocephalus</i>	Anostomidae	1.3Gb	SRR7263034
<i>Moenkhausia sanctaefilomenae</i>	Characidae	1.2Gb	SRR5839692
<i>Piaractus mesopotamicus</i>	Serrasalmidae	1.1Gb	SRR11676685
<i>Prochilodus lineatus</i>	Prochilodontidae	1.6Gb	SRR11676686
<i>Psalidodon paranae</i>	Characidae	1.8Gb	SRR5461470

3.2 - Identificação do gene snRNA U4 e alinhamento das sequências semelhantes para identificação de possíveis variantes

Como já mencionado, os genes snDNA Us estão organizados em repetições em tandem na maioria dos genomas eucarióticos (Silva et al. 2015; Ponzio et al. 2018; Azambuja et al. 2022).

Nesse sentido, uma *pipeline* iterativa foi desenvolvida e aplicada para a caracterização das unidades do gene snDNA U4 repetidas em tandem em 11 espécies de peixes não-modelo da ordem Characiformes (Figura 1).

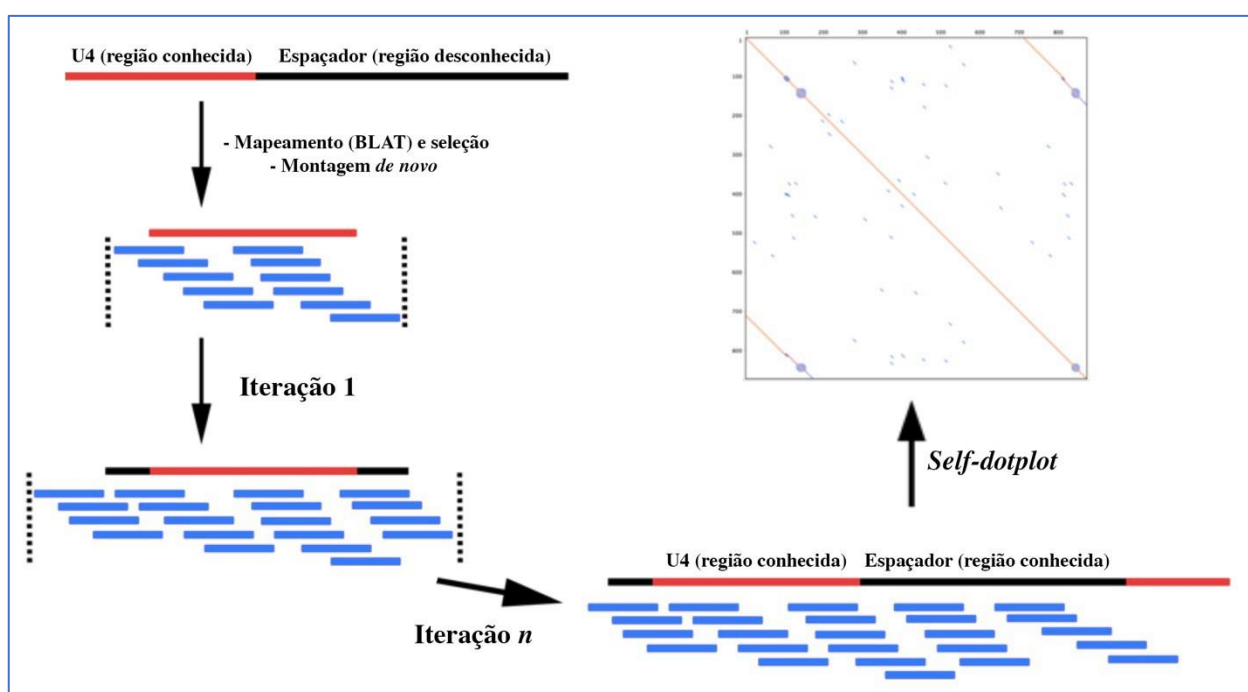


Figura 1: fluxo de etapas consecutivas (iterações) que foram realizadas até a caracterização das unidades completas de snDNA U4 repetidas em tandem das espécies-alvo. Este processo é global e deve funcionar para diferentes famílias multigênicas que se organizam em tandem e possuem homogeneidade entre as repetições.

Primariamente, todas as bibliotecas genômicas utilizadas no presente trabalho foram preparadas, removendo os adaptadores de sequenciamento, passando por um filtro de qualidade (LEADING:3; TRAILING:3; SLIDINGWINDOW:4:15; MINLEN:99; CROP: 100) e padronizando os tamanhos dos *reads* em 2 x 100bp, usando o *software* Trimmomatic (Bolger et

al. 2014). Posteriormente, procedemos à montagem das unidades repetitivas, conforme o descrito a seguir: *i*) usamos uma sequência do gene snDNA U4 humano (*Rfam accession number*: URS00008FED2D_9606) como referência para as análises posteriores em todas as espécies; *ii*) utilizamos o *software* BLAT (Kent 2002) para alinhar *reads* com similaridade ao gene U4 e selecionar estes *reads* com um *script* personalizado (https://github.com/fjruiizuano/ngsprotocols/blob/master/mapping_blat_gs.py); *iii*) Posteriormente, baixamos a seleção de *reads* pareados e realizamos a montagem *de novo* da unidade do snDNA U4 no *software* Geneious 8, usando a opção '*only use paired-end reads*'; *iv*) Avaliar se os *contigs* resultantes possuem o gene U4, bem como se ele representa uma unidade monomérica inteira, por meio de inspeção de um *self-dotplot*. Caso a unidade monomérica não tenha sido caracterizada completamente, recomeçar o processo a partir da etapa *i*, usando o maior *contig* contendo o gene U4 como a referência para a etapa *ii*. Repetir *n* iterações até que o *self-dotplot* revele que a caracterização da unidade completa foi realizada.

3.3 Identificação de SNPs (*Single Nucleotide Polimorphisms*)

Após a caracterização da sequência consenso do snDNA U4 de cada espécie, procedemos com a identificação de SNPs ao longo de cada unidade em cada espécie. Para isso, trabalhando de forma independente em cada espécie, fizemos uma pré-seleção de *reads* com similaridade à unidade completa do gene snDNA U4 utilizando o *software* bowtie2 (Langmead e Salzberg 2012) para mapear *reads* contra um dímero de snDNA U4. Posteriormente, enviamos os arquivos de mapeamento (*.bam) para o *software* Geneious e utilizamos a opção '*map to reference*' contra a unidade do snDNA U4 circularizada (*Minimum coverage*: 10; *Minimum Variant Frequency*: 0.25).

3.4 Avaliação de abundância das unidades de snDNA U4

Com os resultados obtidos no item anterior, extraímos as informações gerais de cobertura por posição em cada espécie ao longo das unidades, visando a identificação de possíveis genes snDNA U4 isolados em diferentes regiões genômicas. Para isso, utilizamos um *script* específico (https://github.com/fjruiizuano/ngs-protocols/blob/master/bam_coverage_join.py) e normalizamos cada resultado pelo número de nucleotídeos analisados em cada biblioteca, estimando o número de cópias ao multiplicar pelo tamanho do genoma. Esta análise permite

inferir se existem cópias adicionais expressivas do gene snDNA U4 nas formas i) órfã (ex: genes espalhados e não organizados em tandem); ou ii) na forma de classes distintas (ex: unidades repetidas em tandem, mas com outros espaçadores. Abaixo, exemplos de cobertura regular e irregular que sugerem diferentes cenários de organização para famílias multigênicas repetidas em tandem (Figura 2; da Silva dados não publicados).

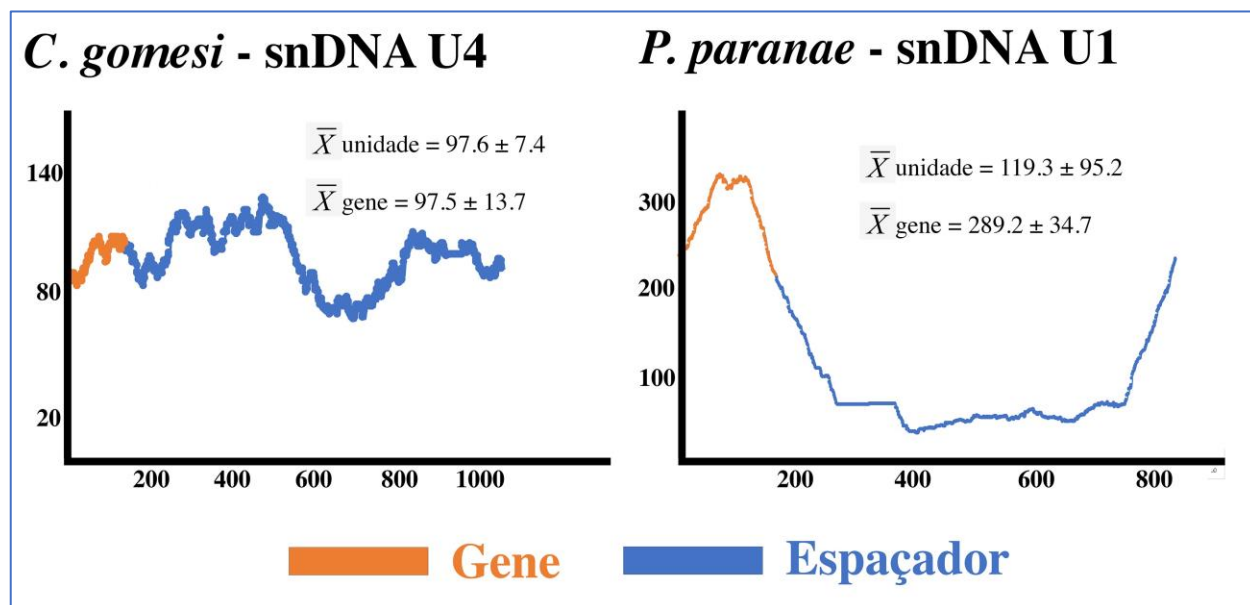


Figura 2: Gráficos de cobertura ilustrando o número de cópias por posição de diferentes genes. O eixo x representa a posição da unidade, e o eixo y representa o número de cópias estimadas em cada espécie. À esquerda, gráfico obtido em snDNA U4 em *C. gomesi*, revelando uma cobertura regular e indicando que não há excesso de cópias adicionais deste gene fora das unidades repetidas em tandem. Por outro lado, à direita, um gráfico para a unidade de snDNA U1 em *P. paranae*, revelando a existência de genes snDNA U1 fora do *cluster* repetido em tandem em análise. Notar a diferença entre as médias do número de cópias observadas para a unidade completa e apenas das regiões gênicas

3.5 - Análises comparativas com genomas completos

Visando verificar se os resultados obtidos na caracterização do gene snDNA U4 a partir de *short-reads* geram unidades completas, bem como realizar análises de divergência intragenômica, nós baixamos o genoma em nível de cromossomo de *Hoplias malabaricus*

(GCA_029633875.1). Posteriormente, isolamos unidades completas dos *arrays* de snDNA U4, utilizando o seguinte protocolo: i) BLAST (Altschul et al. 1990) do gene U4 (*Rfam accession number*: URS00008FED2D_9606) contra os genomas; ii) processamento do *output* para isolar unidades completas (gene + espaçador), utilizando um *script personalizado* (https://github.com/zenirodrigo/repetitives_monomers.sh); iii) alinhamento das unidades isoladas neste processo em conjunto com as unidades das próprias espécies caracterizadas anteriormente utilizando MAFFT (Nakamura et al. 2018).

4. Resultados

4.1 Montagem de unidades de snDNA U4

Após quantidades variadas de múltiplas iterações, a sequência das unidades completas do snDNA U4 (gene + espaçador) repetidas em tandem foi caracterizada para cada uma das espécies. A quantidade de iterações variou entre cada espécie, sendo o mínimo de uma iteração em *H. malabaricus*, até 16 em *C. gomesi*, embora a maioria de espécies tenha exigido em torno de duas ou três iterações para caracterização do gene. Em todas as 11 espécies, a região codificante do gene snDNA U4 se mostrou bastante conservada, apresentando 141 nucleotídeos em todos os indivíduos analisados, com valores de similaridade com o gene snDNA U4 de *Danio rerio* entre 90,43% e 95,74%. Por outro lado, o tamanho e composição das sequências espaçadoras variou bastante entre as espécies, com tamanhos entre 664bp e 1464bp (Figura 3).

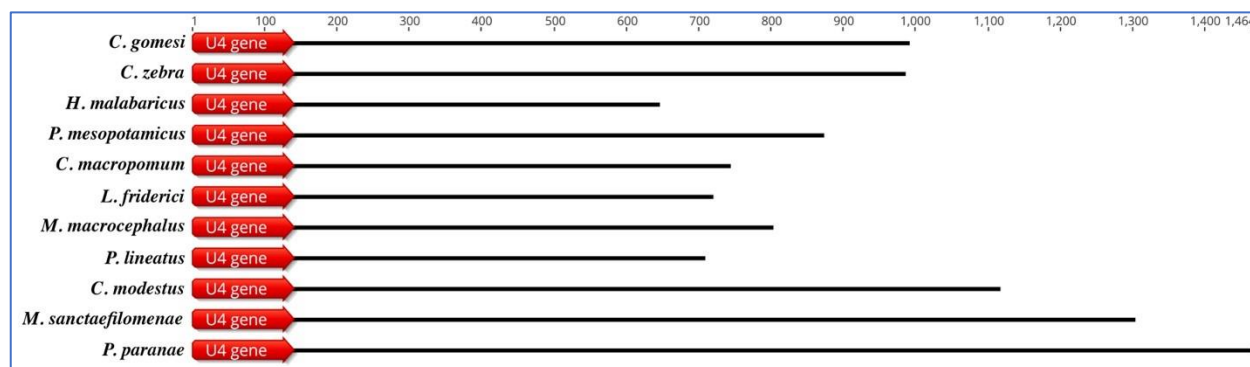


Figura 3: Variação de tamanho da região espaçadora do snDNA U4 entre diferentes espécies de Characiformes.

O alinhamento das unidades completas do snDNA U4 não produziu resultados confiáveis, o que deve estar relacionado à ampla diversificação deste tipo de sequência entre

espécies que divergiram há um longo tempo. No entanto, quando unidades de espécies classificadas em uma mesma família são alinhadas (Ex: *P. mesopotamicus* e *C. macropomum*, *C. zebra* e *C. gomesi*, *L. friderici* e *M. macrocephalus*, e *P. paranae* e *M. sanctaefilomenae*), é possível identificar uma correspondência, também entre as regiões espaçadoras (Figura 4).

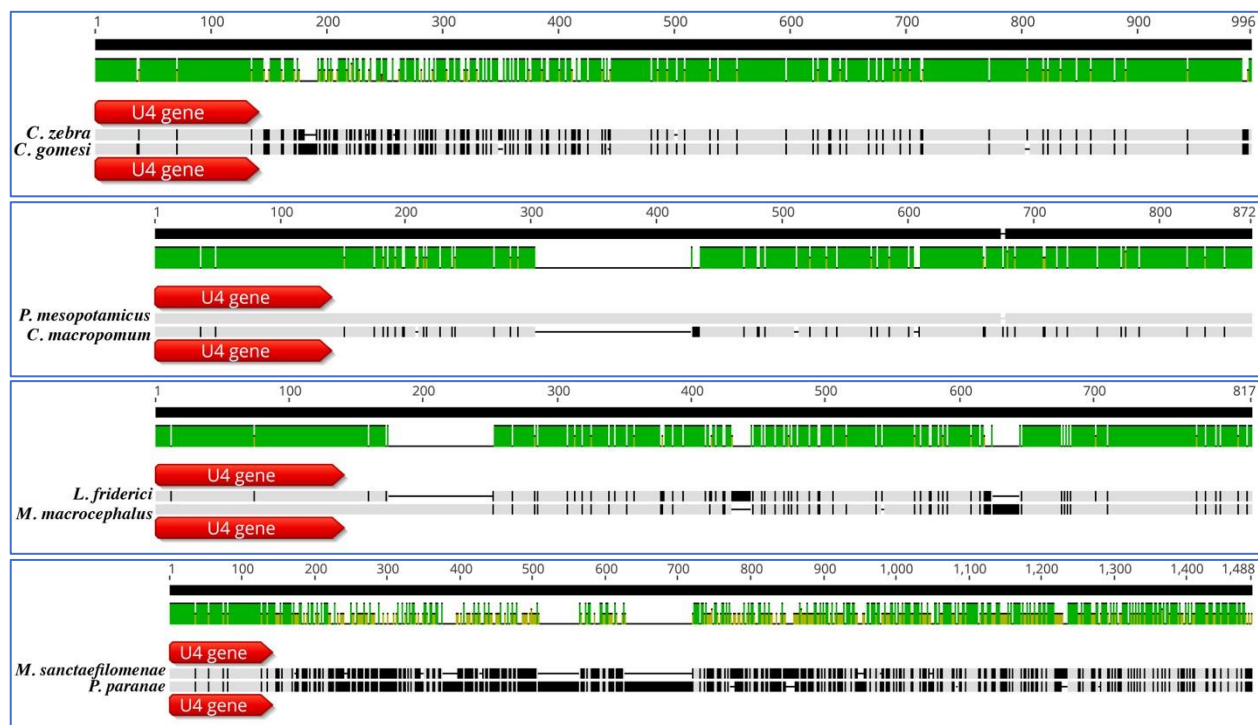


Figura 4: Alinhamentos interespecíficos de unidades completas do snDNA U4. Notar blocos de similaridade nas regiões espaçadoras entre as espécies, indicando uma origem comum destas regiões.

Finalmente, deve-se destacar que as regiões 5' *upstream* à região codificadora apresenta uma certa conservação entre todas as espécies (Figura 5). O que provavelmente está relacionado à regiões promotoras deste gene.

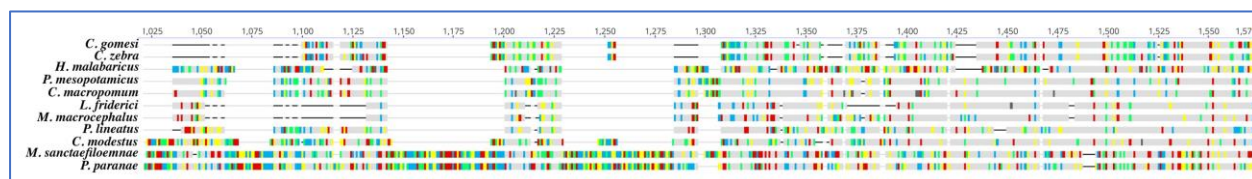


Figura 5: Alinhamentos interespecíficos das regiões 5' *upstream* ao gene snDNA U4, evidenciando blocos conservados.

4.2 Identificação e análise de SNPs

A análise intragenômica de SNPs demonstrou maior variação nos sítios espaçadores em relação à região codificadora do gene snDNA U4. Das 11 espécies analisadas, 7 não apresentaram nenhum SNP na região gênica (*C. macropomum*, *C. modestus*, *L. friderici*, *M. macrocephalus*, *M. sanctaefilomenae*, *P. mesopotamicus*, *P. paranae*), enquanto *C. zebra* apresentou um SNP, *C. gomesi* e *H. malabaricus* apresentaram dois, e a espécie *P. lineatus* se destacou das demais, apresentando seis SNPs na região gênica, todos na primeira metade da porção inicial do gene, com 4 deles nas primeiras 17 bases (Tabela 2).

Quanto à região espaçadora, a espécie *M. macrocephalus* foi a que menos apresentou variações, com apenas três SNPs nessa região, enquanto as espécies *P. lineatus*, *L. friderici* e *H. malabaricus* apresentaram o maior número de variações nos espaçadores, com 38, 22 e 21 SNPs, respectivamente, sendo *P. lineatus* a espécie que exibiu a maior quantidade de SNPs tanto na região gênica, quanto na região espaçadora.

Tabela 2: Tamanho das regiões U4 e espaçadoras, contendo a quantidade e porcentagem de snps em relação ao tamanho da sequência, para cada espécie.

Espécies	U4		Espaçador	
	Bases	SNPs	Bases	SNPs
<i>C. gomesi</i>	141	2	849	12
<i>C. zebra</i>	141	1	843	8
<i>C. macropomum</i>	141	-	602	13
<i>C. modestus</i>	141	-	975	11
<i>H. malabaricus</i>	141	2	503	21
<i>L. friderici</i>	141	-	578	22
<i>M. macrocephalus</i>	141	-	661	3
<i>M. sanctaefilomenae</i>	141	-	1160	14
<i>P. mesopotamicus</i>	141	-	730	7
<i>P. lineatus</i>	141	6	566	38
<i>P. paranae</i>	141	-	1323	11

4.3 Análise de Cobertura

As análises de cobertura revelaram uma distribuição homogênea das sequências, indicando que unidades órfãs, ou classes adicionais repetidas em tandem, não estão amplamente distribuídas ao longo dos genomas das espécies analisadas aqui (Figura 6, Tabela 3).

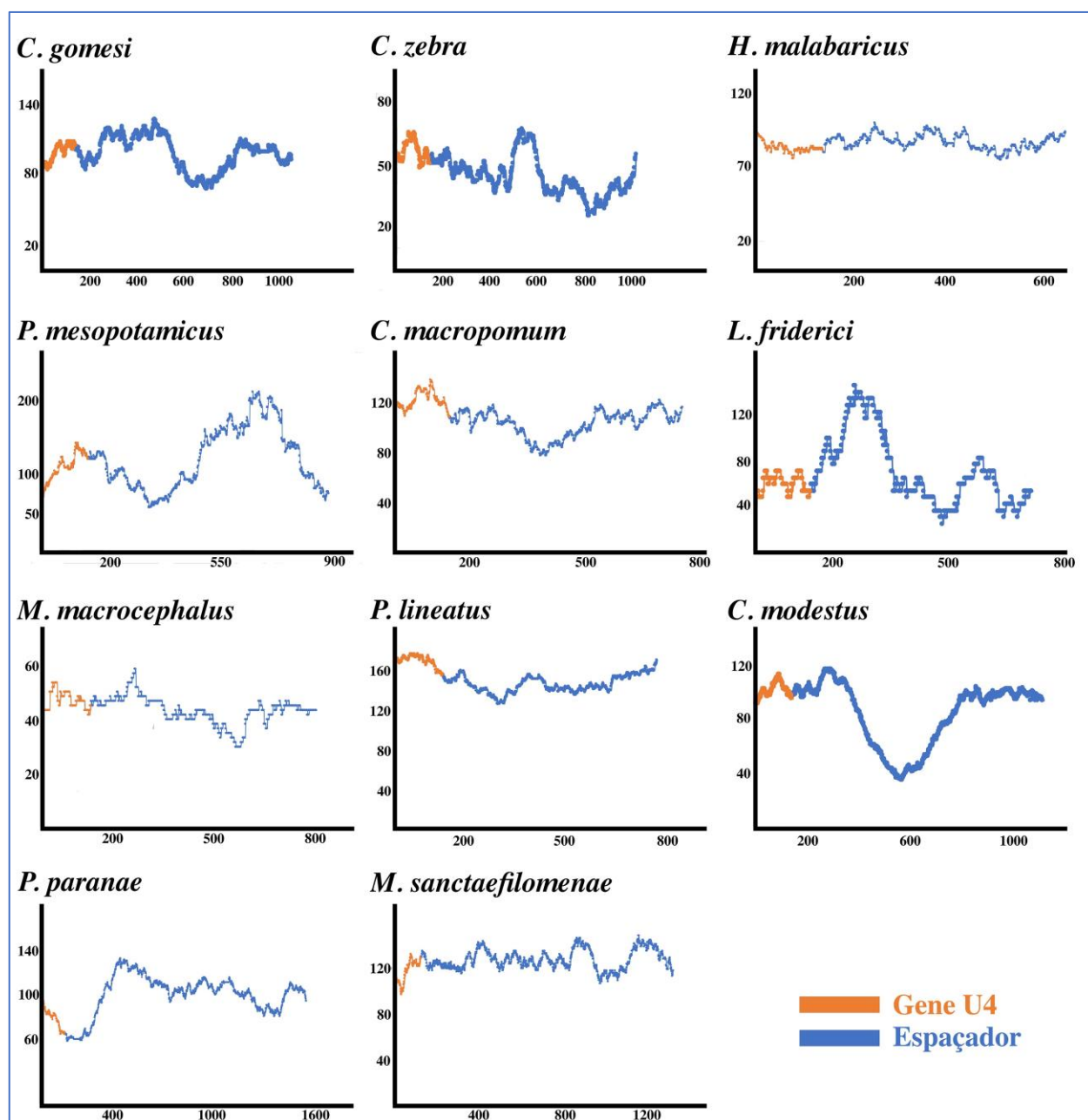


Figura 6: Gráficos de cobertura obtidos para unidades de snDNA U4. O eixo x representa a posição da unidade, e o eixo y representa o número de cópias estimadas em cada espécie

Tabela 3: Valores médios da cobertura da região gênica e espaçadora do snDNA U4 em diferentes espécies de Characiformes.

Espécie	Média – cópias U4	Média – cópias espaçadores
<i>C. gomesi</i>	97.6 ± 7.4	97.5 ± 13.8
<i>C. zebra</i>	56.9 ± 5.0	45.9 ± 9.9
<i>C. macropomum</i>	121.8 ± 7.1	106.5 ± 12.4
<i>C. modestus</i>	101.9 ± 22.9	86.2 ± 5.5
<i>H. malabaricus</i>	82.2 ± 3.2	85.2 ± 4.8
<i>L. friderici</i>	58.7 ± 7.9	67.9 ± 28.7
<i>M. macrocephalus</i>	47.1 ± 3.2	43.9 ± 4.9
<i>M. sanctaefilomenae</i>	117.2 ± 9.4	126.3 ± 9.1
<i>P. mesopotamicus</i>	109.4 ± 16.9	117.8 ± 38.2
<i>P. lineatus</i>	171.2 ± 6.0	152.1 ± 12.3
<i>P. paranae</i>	80.4 ± 9.4	98.8 ± 18.2

4.4 Análise comparativa com genomas completos

A busca por unidades completas de snDNA U4 no genoma de *H. malabaricus* resultou no isolamento de 54 unidades repetidas em tandem a partir do super *scaffold* SUPER_4 (cromossomo 4). Conforme esperado pela teoria, as unidades completas do snDNA U4 são homogêneas entre si, apresentando um baixo nível de diversificação intragenômica (Tabela 4), tanto dos genes, quanto dos espaçadores, e o número de cópias isoladas é próximo à estimativa média de 82 cópias de snDNA U4 obtida a partir de *short-reads* desta espécie (Figura xx – coverage, Tabela xx).

Tabela 4: Valores de diversidade intragenômica a partir das unidades isoladas do genoma de *H. malabaricus*. Notar a baixa diversidade nucleotídica considerando tanto as unidades completas (genes e espaçadores), quanto independentes

	Gene + espaçador	Gene	Espaçador
Div. nucleotídica (π)	0.01879	0.01803	0.019
Haplótipos	48	32	46

5. Discussão

Estudos sobre evolução de famílias multigênicas tiveram início há décadas (revisão em Rebordinos et al. 2013). Em peixes neotropicais, Martins e Galetti Jr. (2001) compilaram uma série de dados evidenciando a existência de, ao menos, duas classes de DNAr 5S. Na ocasião, dados de *Southern blot* e clonagem de fragmentos eram fundamentais para desvendar estas pequenas questões. Ou seja, a simples caracterização de famílias multigênicas era bastante laboriosa e gerava dados limitados ao número de clones sequenciados em cada experimento (Martins e Galetti Jr. 2000; Martins e Wasko 2004; Vierna et al. 2009; Freire et al. 2010; Perina et al. 2011; Pinhal et al. 2011; Scacchetti et al. 2015).

Os genes snDNA U têm sido amplamente estudados em eucariotos, sob aspectos funcionais e estruturais, principalmente pelo seu papel central no mecanismo de *splicing* (Newman e Norman 1992; Vierna et al. 2013). No entanto, estas características gerais são basicamente limitadas às regiões codificantes destes genes, principalmente a partir de espécies com genomas montados e sequenciados (Marz et al. 2008; Cabral-de-Mello et al. 2012). Nesse contexto, os dados apresentados aqui abrem novas perspectivas para estudos de organização genômica e evolução destes genes, uma vez que o método apresentado aqui permite a caracterização de unidades completas (região gênica e espaçadora) em espécies não-modelo a partir de dados de sequenciamento de *short-reads* de baixa cobertura.

Até a presente data, alguns trabalhos descreveram formas alternativas de reconstrução e análise de unidades de famílias multigênicas. Marti et al. (2021), por exemplo, utilizaram o *software* NOVOPlasty (Dierckxsens et al. 2017) para reconstruir diferentes *clusters* em espécies de gafanhotos. Por ser uma *pipeline* desenvolvida para reconstrução de genomas organelares (ex: mitocôndrias), esta é uma estratégia interessante, mas que não permitiu a reconstrução de unidades completas no presente trabalho (dados não mostrados). A *pipeline* RepeatExplorer também permite a reconstrução de unidades completas de famílias multigênicas (Palacios-Gimenez et al. 2020), mas é fundamental que as sequências apresentem abundância elevada, o que não é o caso de diversos genes snDNA Us. Desta forma, o método apresentado aqui representa uma alternativa viável para a caracterização de famílias multigênicas, inclusive aquelas que se dividem em classes distintas, como é o caso do snDNA U1 em peixes da família Characidae (da Silva *in prep.*). Em termos de replicabilidade, o método bioinformático descrito

aqui permite o isolamento das mesmas classes de famílias repetitivas estudadas com estratégias de clonagem e sequenciamento de Sanger (da Silva *in prep.*).

Conforme predito pela teoria em sistemas que evoluem em concerto, as unidades de snDNA U4 mostraram baixa variação intragenômica, com SNPs distribuídos de maneira esparsa ao longo das regiões codificadoras e espaçadoras. Além disso, a diferenciação das regiões espaçadoras conforme a proximidade filogenética das espécies é uma característica marcante e revela que estas regiões não devem desempenhar um papel fundamental no funcionamento destes genes, exceto pela região 5' *upstream*.

6. Referências

- ALMEIDA-TOLEDO, L. F. Cytogenetic markers in neotropical freshwater fishes. *In*: MALABARBA, L. R.; REIS, R. E.; VARI, R. P.; LUCENA, Z. M. S.; LUCENA, C. A. **S. Phylogeny and Classification of Neotropical Fishes**. Porto Alegre: Edipucrs, 1998, p. 583-588.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.
- ANJOS, A. et al. Insights into chromosomal evolution of Cicadomorpha using fluorochrome staining and mapping 18S rRNA and H3 histone genes. **Journal of Zoology and Systematic Evolution**, v. 57, p. 314-322, 2019.
- ANJOS, A. et al. U1 snDNA clusters in grasshoppers: chromosomal dynamics and genomic organization. **Heredity**, v. 114, n. 2, p. 207-219, 2015.
- ANLEITNER, J. E.; HAYMER, D. S. Y enriched and Y specific DNA sequences from the genome of the Mediterranean fruit fly, *Ceratitis capitata*. **Chromosoma**, v. 101, p. 271-278, 1992.
- ARTONI, et al. An uncommon condition for a sex chromosome system in Characidae fish. Distribution and differentiation of ZZ/ZW system in *Triportheus*. **Chromosome Research**, v. 9, p. 449-456, 2001.
- AZAMBUJA, M. et al. Major and minor U small nuclear RNAs genes characterization in a neotropical fish genome: Chromosomal remodeling and repeat units dispersion in Parodontidae. **Gene**, v. 826, p. 146459, 2022.
- BARDELLA, V. B.; CABRAL-DE-MELLO, D. C. Uncovering the molecular organization of unusual highly scattered 5S rDNA: the case of *Chariesterus armatus* (Heteroptera). **Gene**, v. 646, p. 153-158, 2018.
- BERTOLLO, L. A. C. et al. A biodiversity approach in the Neotropical fish *Hoplias malabaricus*. Karyotypic survey, geographic distribution of cytotypes and cytotaxonomic considerations. **Chromosome Research**, v. 8, pg 603-613, 2000.
- BISCOTTI, M. A. et al. Repetitive DNA in eukaryotic genomes. **Chromosome Research**, v. 23, n. 3, p. 415-420, 2015.
- BOLGER, M. A. et al. Trimmomatic: a flexible trimmer for illumine sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014.

- BRINGMANN P.; LÜHRMANN, R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. **The EMBO Journal**, v. 5, n. 13, p. 3509-3516.
- BURGE, C. B.; PADGETT, R. A.; SHARP, P. A. Evolutionary fates and origins of U12-type introns. **Molecular Cell**, v. 2, n. 6, p. 773-785, 1998.
- CABRAL-DE-MELLO, D. C.; MOURA, R. C. MARTINS, C.; Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. **Heredity**, v. 104, p. 393-400, 2010.
- CABRAL-DE-MELLO, et al. Cytogenetic mapping of 5S and 18S rRNAs and H3 histone genes in 4 ancient Proscopiidae grasshopper species: contribution to understanding the evolutionary dynamics of multigene families. **Cytogenetic and Genome Research**. n. 132, p. 89-93, 2011a.
- CABRAL-DE-MELLO, et al. Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. **Genetica**, v. 139, p. 921-931, 2012.
- CABRAL-DE-MELLO, et al. Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. **Chromosome Research**, v. 20, n. 2, p. 279-292, 2012.
- CABRERO, et al. Comparative FISH analysis in five species of Eyprepocnemidine grasshoppers. **Heredity**, v. 90, p. 377-381, 2003.
- CABRERO, J. et al. Chromosome mapping of H3 and H4 histone gene clusters in 35 species of acridid grasshoppers. **Chromosome Research**, v. 17, p. 397-404, 2009.
- CABRERO, J.; CAMACHO, J. P. Location and expression of ribosomal RNA genes in grasshoppers: abundance of silent and cryptic loci. **Chromosome**, v. 16, n. 4, p. 595-607, 2008.
- CARADONNA, F. et al. Chromosomal localization and molecular characterization of three different 5S ribosomal DNA clusters in the sea urchin *Paracentrotus lividus*. **Genome**, v. 50, p. 867-870, 2007.
- CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v. 371, p. 215-220, 1994.

- CIOFFI, M. B. et al. Differentiation of the XY sex chromosomes in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae). Unusual accumulation of repetitive sequences on the X chromosome. **Sexual Development**, v. 4, n. 3, p. 176–185, 2010.
- CROS, I.; REBORDINHOS, L.; 5S rDNA and U2 snRNA are linked in the genome of *Crassostrea angulata* and *Crassostrea gigas* oysters: does the (CT)_n – (GA)_n microsatellite stabilize this novel linkage of large tandem arrays? **Genome**, v. 48, p. 1116–1119, 2005.
- DATSON, P. M.; MURRAY, B. G. Ribosomal DNA locus evolution in *Nemesia*: transposition rather than structural rearrangement as the key mechanism? **Chromosome Research**, v. 14, p. 845–857, 2006.
- DEGRANDI, T. M. et al. The distribution of 45S rDNA sites in bird chromosomes suggests multiple evolutionary histories. **Genetic and Molecular Biology**, v. 2, p. 43, 2020.
- DIERCKXSENS, N.; MARDULYN, P.; SMITHS, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. **Nucleic Acids Research**, v. 45, n. 4, p. e18, 2017.
- DOVER, G. Molecular Drive. **Trends in Genetics**, v. 18, n. 11, p. 587–589, 2002.
- DOVER, G. Molecular drive: a cohesive mode of species evolution. **Nature**, v. 299, p. 111–117, 1982.
- DROUING, G.; MONIZ DE SÁ, M. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. **Molecular Biology and Evolution**, v. 12, p. 481–493, 1995.
- EICKBUSH, T. H.; EICKBUSH, D. G. Finely orchestrated movements: evolution of the ribosomal RNA genes. **Genetics**, v. 175, p. 477–485, 2007.
- EIRÍN-LÓPEZ, J. M. et al. The birth-and-death evolution of multigene families revisited. **Genome Dynamics**, v. 7, p. 170–196, 2012.
- ENGEL, J. D.; DODGSON, J. B. Histone genes are clustered but not tandemly repeated in the chicken genome. **Proceedings of the National Academy of Sciences**, v. 78, n. 5, p. 2856–2860, 1981.
- ESCHMEYER, W. N. R.; FRICKE, R. **ESCHMEYER'S CATALOG OF FISHES: GENERA, SPECIES**, van der Laan (eds), 2023.
<https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>.

Versão eletrônica acessada em 30/07/2023.

- FERNANDEZ, M. et al. 5S rDNA genome regions of *Lens* species. **Genome**, v. 42, p. 937-942, 2005.
- FERRETTI, A. et al. How dynamic could be the 45S rDNA cistron? An intriguing variability in a grasshopper species revealed by integration of chromosomal and genomic data. **Chromosoma**, v. 128, p. 165-175, 2019.
- FORESTI, F.; ALMEIDA-TOLEDO, L. F.; TOLEDO-FILHO, A. S. Polymorphic nature of nucleolar organizer regions in fishes. **Cytogenetics and Cell Genetics**, v. 31, p. 137-144, 1981.
- FREIRE, R. et al. Evolutionary dynamics of the 5S rDNA gene family in the mussel *Mytilus*: mixed effects of birth-and-death and concerted evolution. **Journal of Molecular Evolution**, v. 70, p. 413-426, 2010.
- GARCIA-SOUTO, D. et al. Molecular cytogenetic analysis of the European hake *Merluccius merluccius* (Merlucciidae, Gadiformes): U1 and U2 snRNA gene clusters map to the same location. **PLoS ONE**, v. 10, n. 12, p. e0146150, 2015.
- GARCIA-SOUTO, D. S. et al. Karyotype characterization of nine periwinkle species (Gastropoda, Littorinidae). **Genes**, v. 9, p. 517, 2018.
- GARRIDO-RAMOS, M. A. Satellite DNA in plants: More than just Rubbish. **Cytogenetic and Genome Research**, v. 146, n. 2, p. 153–170, 2015.
- GARRIDO-RAMOS, M. A. Satellite DNA: An evolving topic. **Genes**, v. 8, n. 9, 2017.
- GOUBERT, C. et al. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). **Genome Biology and Evolution**, v. 7, n. 4, p. 1192-1205, 2015.
- HASHIMOTO, D. T. et al. Chromosome mapping of H1 histone and 5S rRNA gene clusters in three species of *Astyanax* (Teleostei, Characiformes). **Cytogenetic and Genome Research**, v. 134, n. 1, p. 64-71, 2011.
- HUGHES, A. L.; NEI, M. Maintenance of MHC polymorphism. **Science**, v. 355, p. 402-403, 1992.
- INNAN, H. KONDRASHOV, F. The evolution of gene duplications: Classifying and distinguishing between models. **Nature Reviews Genetics**, v. 11, p. 97-108, 2010.

- JENUWEIN, T.; ALLIS, C. D. Translating the histone code. **Science**, v. 293, p. 1074-1080, 2001.
- KENT, W. J. BLAT —The BLAST -Like Alignment Tool. **Genome Research**, v. 12, n. 4, p. 656– 664, 2002.
- KINDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes. **Genetica**, v. 115, p. 49-63, 2002.
- KRAEMER, C.; SCHMIDT, E. R. The sex-determining region of *Chironomus thummi* is associated with highly repetitive DNA and transposable elements. **Chromosoma**, v. 102, p. 553-562, 1993.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie2. **Nature Methods**, v, 9, n.4, p. 357–359, 2012.
- LERNER, M. R. et al. Are snRNPs involved in splicing? **Nature**, v. 283, n. 5743, p. 220-224, 1980.
- LIFTON, R. P. et al. The organization of the histone genes in *Drosophila melanogaster*: Functional and evolutionary implications. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 42, p. 1047-1051, 1977.
- LONG, E. O.; DAWID, I. B.; Repeated genes in eukaryotes. **Annual Review of Biochemistry**, v. 49, p. 727-767, 1980.
- LÓPEZ, M. D.; ROSENBLAD, M. A.; SAMUELSSON, T. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. **Nucleic Acids Research**, v. 36, P. 3001-3010, 2008.
- LUNDBERG, J. G. et al. So many fishes, so little time: an overview of recent ichthyological discovery in continental waters. **Annals of the Missouri Botanical Garden**, v. 87, n. 1, p. 26-62, 2000.
- MANCHADO, M. et al. Molecular characterization and chromosomal mapping of the 5S rRNA gene in *Solea senegalensis*: a new linkage to the U1, U2, and U5 small nuclear RNA genes. **Genome**, v. 49, p. 79-86, 2006.
- MARTI, E. et al. Cytogenomic analysis unveils molecular evolution and recurrent chromosomal rearrangements shaping the multigene families on *Schistocerca* grasshopper genomes. **Evolution**, v. 75, n. 8, p. 2027-2041. 2021.

- MARTINS, C. Chromosomes and repetitive DNAs: a contribution to the knowledge of fish genome. *In*: PISANO, E.; OZOUF-COSTAZ, C.; FORESTI, F.; KAPOOR, B. G. **Fish Citogenetics**. Enfield, New Hampshire: Science Publisher, 2007, p. 421-453.
- MARTINS, C.; GALETTI JR, P. M. Conservative distribution of 5S rDNA loci in *Schizodon* (Pisces, Anostomidae) chromosomes. **Chromosome Research**, v. 8, p. 353-355, 2000.
- MARTINS, C.; GALETTI Jr P. M. Two rDNA arrays in Neotropical fish species: is it a general rule for fishes? **Genetica**, v. 111, p. 439–446, 2001.
- MARTINS, C; WASKO, A. P. Organization and evolution of 5S ribosomal DNA in the fish genome. *In*: WILLIAMS, C. R. (Ed.). **Focus on Genome Research**. New York: Nova Science Publishers, 2004, p. 335-363.
- MARZ, M.; KIRSTEN, T.; STADLER, P. F. Evolution of spliceosomal snRNA genes in metazoan animals. **Journal of Molecular Evolution**, v. 67, p. 594-607, 2008.
- MAZZONENI, S. et al. Evolutionary insight on localization of 18S, 28S rDNA genes on homologous chromosomes in Primates genomes. **Comparative Citogenetics**, v. 12, p. 2740, 2018.
- MERLO, M. A. et al. Evidence for 5S rDNA horizontal transfer in the toadfish *Halobatrachus didactylus* (Schneider, 1801) based on the analysis of three multigene families. **BMC Ecology and Evolution**, v. 12, p. 201, 2012.
- MESSIER, W.; LI, S. H.; STEWART, C. B. The birth of microsatellites. **Nature**, v. 381, p. 483, 1996.
- MESTRINER, C. A. et al. Structural and functional evidence that a B chromosome in the characidae fish *Astyanax scabripinnis* is an isochromosome. **Heredity**, v. 85, p. 1-9, 2000.
- MEWES, H. W. et al. Overview of the yeast genome. **Nature**, v. 387, n. 6632, p. 7-65, 1997.
- MOREIRA, C. R. et al. Relações filogenéticas na ordem Characiformes 1274 (Teleostei: Ostariophysi). **Unpublished Ph. D. Dissertation, Universidade de São Paulo**, v. 468, 2007.
- MOUNT, S. M. et al. Spliceosomal small nuclear RNA genes in 11 insect genomes. **RNA**, v. 13, p. 5-14, 2007.
- NAKAMURA, T. et al. Parallelization of MAFFT for large-scale multiple sequence alignments. **Bioinformatics**, v. 34, n. 14, p. 2490-2492, 2018.

- NEI, M.; ROONEY, A. P. Concerted and birth-and-death evolution of multigene families. **Annual Review of Genetics**, v. 39, p. 121-152, 2005.
- NEWMAN, A. J.; NORMAN, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. **Cell**, v. 68, n. 4, p. 743-754, 1992.
- NGUYEN, P. K.; SAHARA, A. Y.; MAREC, F. Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). **Genetica**, v. 138, p. 343-354, 2010.
- NILSEN, T. W. The spliceosome: the most complex macromolecular machine in the cell? **BioEssays**, v. 25, p. 1147-1149, 2003.
- NOVÁK, P. et al. Genome-Wide analysis of repeat diversity across the family Musaceae. **Plos One**, v. 9, n. 6, p. e98918, 2014.
- NOVÁK, P. et al. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792–793, 2013.
- NOVÁK, P. et al. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Research**, v. 45, n. 12, 2017.
- OLIVEIRA, C.; FORESTI, F.; HILSDORF, A.W.S. Genetics of Neotropical fish: from chromosomes to populations. **Fish Physiology and Biochemistry**, v. 35, p. 81-100, 2009.
- PALACIOS-GIMENEZ, et al. Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats. **BMC Biology**, v. 18, n. 1, p. 1-21, 2020.
- PALACIOS-GIMENEZ, et al. Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. **Genome Biology and Evolution**, v. 12, n. 3, p. 88–102, 2020.
- PATEL, A. A.; STEITZ, J. A. Splicing double: insights from the second spliceosome. **Molecular and Cell Biology**, v. 4, n. 12, p. 960-670, 2003.
- PELLICCIA, et al. 5S ribosomal and U1 small nuclear RNA genes: a new linkage type in the genome of a crustacean that has three different tandemly repeated units containing 5S ribosomal DNA sequences. **Genome**, v. 44, p. 331-335, 2001.
- PERINA, et al. Molecular organization and phylogenetic analysis of 5S rDNA in crustaceans of the genus *Pollicipes* reveal birth- and death evolution and strong purifying selection. **BMC Ecology and Evolution**, v. 11, p. 304, 2011.

- PETERSON, R. C.; DOERING, J. L.; BROWN, D. D. Characterization of two *Xenopus* somatic 5S-DNAs and one minor oocyte-specific 5S-DNA. **Cell**, v. 20, p. 131-141, 1980.
- PINHAL, et al. The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. **BMC Ecology and Evolution**, v. 11, p. 151, 2011.
- PISCOR, D.; FERNANDES, C.; & PARIDE-MALTEMPI, P. P. Conserved number of U2 snDNA sites in *Piabina argentea*, *Piabarchus stramineus* and two *Bryconamericus* species (Characidae, Stevardiinae). **Neotropical Ichthyology**, v. 16, n. 1, p. e170066, 2008.
- PLOHL, M. et al. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. **Gene**, v. 409, p. 72-82, 2008.
- PONZIO, J. C.; PISCOR, D.; PARISE-MALTEMPI, P. P. Chromosomal locations of U2 snDNA clusters in *Megaleporinus*, *Leporinus* and *Schizodon* (Characiformes: Anostomidae). **Biologia (Bratislava)**, v. 73, n. 3, p. 295-298, 2018.
- REBORDINOS, L.; CROSS, I.; MERLO, A. High evolutionary dynamism in 5s rDNA of fish: State of the art. **Cytogenetic and Genome Research**, v. 141, n. 2-3, p. 103-113, 2013.
- REIS, R. E. et al. Fish biodiversity and conservation in South America. **Journal of Fish Biology**, v. 89, n. 1, p. 12-47, 2016.
- ROONEY, A. P.; WARD, T. J. Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. **Proceedings of the National Academy of Sciences**, V. 104, N. 14, P. 5084-5089, 2005.
- RUBERTI, et al. Characterization of histone genes isolated from *Xenopus laevis* and *Xenopus tropicalis* genomic libraries. **Nucleic Acids Research**, v. 10, p. 1544-1550, 1982.
- RUIZ-RUANO, F. J. et al. High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports**, v. 6, p. 1–14, 2016.
- RUSSEL, A. G.; CHARETTE, D. F.; SPENCER, M. W. An early evolutionary origin for the minor spliceosome. **Nature**, v. 443, v. 7113, p. 863-866, 2006.
- SCACCHETTI, P. C. et al. Repetitive DNA sequences and evolution of ZZ/ZW sex chromosomes in *Characidium* (Teleostei: Characiformes). **PLoS ONE**, v. 10, n. 9, p. e0137231, 2015.

- SHETH, N. et al. Comprehensive splice-site analysis using comparative genomics. **Nucleic Acids Research**, v. 34, p. 3955-3967, 2006.
- SHI, J. et al. Widespread Gene Conversion in Centromere Cores. **PLoS Biology**, v. 8, n. 3, p. e1000327, 2010.
- SHI, Y. Mechanistic insights into precursors messenger RNA splicing by the spliceosome. **Nature Reviews Molecular and Cell Biology**, v. 18, n. 11, p. 655-670, 2017.
- SHIPPEN-LENTZ, D. R.; VEZZA, A. C. The three 5S rRNA genes from the human malaria parasite *Plasmodium falciparum* are linked. **Molecular and Biochemical Parasitology**, v. 27, p. 263-273, 1988.
- SILVA, D. M. Z. A. et al. Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): dispersive location, association and co-localization in the genome. **Genetica** v. 141, p. 329-336, 2013.
- SILVA, D. M. Z. A. et al. Chromosomal mapping of repetitive DNA sequences in five species of *Astyanax* (Characiformes, Characidae) reveals independent location of U1 and U2 snRNA sites and association of U1 snRNA and 5S rDNA. **Cytogenetic and Genome Research**, v. 146, n. 2, p. 144-152, 2015.
- SILVA, D. M. Z. A. et al. High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. **Scientific Reports**, v. 7, n. 1, p. 1–12, 2017.
- SMITH, G. P. Evolution of repeated DNA sequences by unequal crossover. **Science**, v. 191, p. 528-535, 1976.
- SOARES, L. B. et al. Molecular cytogenetic analysis and the establishment of a cell culture in the fish species *Hollandichthys multifasciatus* (Eigenmann & Norris, 1900) (Characiformes, Characidae). **Genetics and Molecular Biology**, v. 44, n. 2, p. e20200260, 2021.
- SUMNER, A. T. **Chromosomes – organization and function**. Londres: Blackwell Publishing Company, 2003.
- SUPIWONG, W. et al. Karyotype and cytogenetic mapping of 9 classes of repetitive DNAs in the genome of the naked catfish *Mystus bocourti* (Siluriformes, Bagridae). **Molecular Cytogenetics**, v. 6, n. 1, p. 51, 2013.
- TURUNEN, J. J. et al. The significant other: splicing by the minor spliceosome. **Wiley Interdisciplinary Reviews RNA**, v. 4, n. 1, p. 61-79, 2013.

- UGARKOVIC, D.; PLOHL, M.; Variation in satellite DNA profiles—causes and effects. **The EMBO Journal**, v. 21, n. 21, p. 5955-5959, 2002.
- UTSUNOMIA, et al. Comparative chromosome mapping of U2 snRNA and 5S rRNA genes in *Gymnotus* species (Gymnotiformes, Gymnotidae): Evolutionary dynamics and sex chromosome linkage in *G. pantanal*. **Cytogenetic and Genome Research**, v. 142, n. 4, p. 286-292, 2014.
- VALADKHAN, S. Young Scientist Award essay winner. Construction of a minimal protein-free spliceosome. **Science**, v. 307, p. 863-864, 2005.
- VAN-HOLDE, K. E. **Chromatin**. 1st ed. New York: Springer-Verlag, 1988.
- VICARI, M. R. et al. Co-localization of repetitive DNAs and silencing of major rRNA genes. A case report of the fish *Astyanax janae*. **Cytogenetic and Genome Research**, v. 122, p. 67-72, 2008.
- VICENTE, V. E. et al. Origin and differentiation of a sex chromosome system in *Parodon hilarii* (Pisces, Parodontidae). Satellite DNA, G-and C-banding. **Genetica**, v. 119, p. 115-120, 2003.
- VIERNA, J. et al. Systematic analysis and evolution of 5S ribosomal DNA in metazoans. **Heredity**, v. 111, n. 5, p. 410-421, 2013.
- VIERNA, J. et al. The linked units of 5S rDNA and U1 snDNA of razor shells (Mollusca: Bivalvia: Pharidae). **Heredity**, v. 107, n. 2, p. 127-142, 2011.
- VIERNA, J.; GONZÁLEZ-TIZÓN, A. N.; MARTÍNEZ, A. Long-term evolution of 5S ribosomal DNA seems to be driven by birth-and-death processes and selection in Ensis razor shells (Mollusca: Bivalvia). **Biochemical Genetics**, v. 47, n. 9-10, p. 635–644, 2009.
- WALSH, J. B. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. **Genetics**, v. 115, p. 553-567, 1987.
- WANG, W. et al. The multigene family of fungal laccases and their expression in the white rot basidiomycete *Flammulina velutipes*. **Gene**, v. 563, n. 2, p 142-149, 2015.
- WEST, S. The increasing functional repertoire of U1 snRNA. **Biochemical Society Transactions**, v. 40, n.4, p. 846-849, 2012.

- WILL, C. L.; LUHRMANN, R. Spliceosome, structure, and function. **Cold Spring Harbor Perspectives in Biology**, v. 3, n. 7, p. a003707, 2011.
- WILLIAMS, S. M.; STROBECK, C. Sister chromatid exchange and the evolution of rDNA spacer length. **Journal of Theoretical Biology**, v. 116, p. 625-636, 1985.
- YANO, R. et al. Comparative genomics of muskmelon reveals a potential role for retrotransposons in the modification of gene expression. **Communications Biology**, v. 3, p. 432, 2020.
- YUAN, Z. et al. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. **BMC Genomics**, v. 19. N. 1, p. 1-10, 2018.
- ZHANG, M. et al. Concerted and birth-and-death evolution of 26S ribosomal DNA in *Camellia* L. **Annals of Botany**, v. 127, p. 63-73, 2021.